Partie I : Statistiques

Les caractéristiques sont des méthodes pour collecter, présenter et analyser des données.

- Exemple : Etude de l'age des étudiants de l'ensaf

+ l'age : le caractère à étudier, variable statistique ou observation.

+ Les étudiants : population, chaque étudiant est un individu.

+ Echantillon : c'est un sous ensemble de la population ($\mathcal{P}$).

- collecter : 19, 20, ..., on les appelle des données brutes (n'ont pas de sens).

→ information = donnée + sens.

+ Connaissance : utilisateur et interprétateur de l'information.

- Card ($\mathcal{P}$) = N, N est appellé l'effective totale.

- modalité : la valeur prise pour notre variable statistiques

on note X = variable statistique

et x_i les modalités

prenons par exemple $N = 10$.

$\mathcal{P} = \{19, 21, 25, 19, 22, 21, 25, 19, 20, 20\}$

$x_1 = 19$, $x_2 = 20$, $x_3 = 21$, $x_4 = 22$, $x_5 = 25$

$n_1 = 3$, $n_2 = 2$, $n_3 = 2$, $n_4 = 1$, $n_5 = 2$.

Scanned with CamScanner
Scanned with CamScanner

- Série statistique (simple)

x_i	19	20	21	22	25
n_i	3	2	2	1	2

n_i Correspond au nombre d'individus qui prend la valeur du caractère x_i. tq : $N = \sum_{i=1}^{n} n_i$

- Série statistique des effectifs croissantes :

x_i	19	20	21	22	25
$n_i \uparrow$	3	5	7	8	10

n_i effectif relatif à la modalité x_i

- La fréquence : la fréquence relative à la modalité n_i est notée par $f_i = \frac{n_i}{N}$

Exp : $f_1 = \frac{n_1}{10} = \frac{3}{10} = 30\%$.

- Interprétation : 30% des étudiants de l'ENSAF ont un âge de 19 ans.

- Série statistiques des fréquences :

$\sum_{i=1}^{n} f_i = 100\%$

x_i	19	20	21	22	25
f_i	30%	20%	20%	10%	20%

f_i Correspond au pourcentage des individus possédant le caractère x_i

- Série statistique des fréquence croissante :

x_i	19	20	21	22	25
$f_i \uparrow$	30%	50%	70%	80%	100%

Scanned with CamScanner
Scanned with CamScanner

- **Différents types de caractères de variable statistiques**

1/ Les caractères qualitatifs :

un caractère qualificatif est un caractère dont les modalités échappent à la mesure. Elle peuvent seulement être constatées.

- Exemple :
  - Sexe de la personne interrogeé, situation familiale
  - Etat du météo à un certain endroit (pluvieux, ensoleillé, ...)

2°/ Les caractères quantitatifs :

Une variable statistique est dite quantitative si ses valeurs sont des nombres sur lequels on peut effectuer des opérations arithmétiques telles que la somme, la moyenne ... (ont un sens).

- Exemple :
  - Taille / poids / salaire ....
  - Rendement / Notes ....
  - PNB

on distinguer entre deux types de variables caractéristiques quantitatives :

- discrètes : si X prend des valeurs naturelles
- Continues : Si X prend des valeurs réelles.

- **Exemple** : X quantitative

Etude de sexe des etudiants de l'ENSAF

$X = \{M, F\}$, $N = 10$ $X = [M, F, F, M, F, F, M, M, F]$

Scanned with CamScanner
Scanned with CamScanner

(1) → M , (2) → F

x_i	1	2
n_i	4	6

x_i	1	2
f_i	40%	60%

3. Continue : Une variable statistique est continue lorsqu'elle peut prendre toutes les valeurs à l'intérieur de son intervalle de variation. Ce nombre de valeurs possibles est toujours infini, il est donc nécessaire, avant de classer les observations, de définir les modalités du caractère en groupant en classes les valeurs possibles de la variable statistique.

- Exemples de variables continues :
  - la taille, le poids, l'âge d'un individu.
  - la durée de chomâge d'une personne.
  - la distance séparant deux points.

- La moyenne (Espérence mathématique)

Soit X V.Q.D.

x_i	x_1	x_2	...	...	x_n
n_i	n_1	n_2	...	...	n_m

$$E(X) = \overline{X} = \frac{1}{N} \sum_{i=1}^{n} x_i n_i$$

$$E(aX+b) = a \cdot E(X) + b.$$

- Démonstration : on pose $Y = aX + b \Rightarrow y_i = ax_i + b$

$$E(Y) = \frac{1}{N} \Sigma\, i y_i = \frac{1}{N} \Sigma\, i (ax_i + b)$$

$$= a \cdot \frac{1}{N} \cdot \Sigma\, n_i x_i + \frac{b}{N} \Sigma\, n_i = a \cdot E(X) + b.$$

Scanned with CamScanner
Scanned with CamScanner

- **La variance :**

$$V(X) = E(X^2) - E(X)^2$$

$$E(X^2) = \frac{1}{N} \cdot \sum_{i=1}^{n} n_i x_i^2$$

$$\begin{cases} E(a) = a. \\ V(a) = 0 \end{cases} \quad a = cst$$

- Mq $V(aX+b) = a^2 . V(X)$.

$$V(aX+b) = E\left((aX+b)^2\right) - \left(E(aX+b)\right)^2$$

$$= \frac{1}{N} \cdot \sum n_i (aX+b)^2 - \left(a . E(X) + b\right)^2$$

$$= a^2 . \frac{1}{N} \cdot \sum n_i x_i^2 + b^2 . \frac{1}{N} n_i + 2ab . \frac{1}{N} \cdot \sum n_i x_i - \dots$$

$$= a^2 . E(X^2) + 2ab\, E(X) + b^2 - a^2 E^2(X) + 2ab\, E(X) + b^2$$

$$= a^2 \left(E(X^2) - E^2(X) \right)$$

$$= a^2 . V(X).$$

- Mq $V(X) = \frac{1}{N} \sum n_i \left(x_i - \overline{x}\right)^2$

$$\left(\frac{1}{N} \cdot \sum n_i \left(x_i - \overline{x}\right)^2 = \frac{1}{N} \cdot \sum n_i \left(x_i^2 - 2x_i\overline{x} + \overline{x}^2\right) \right.$$

$$= \frac{1}{N} \sum n_i x_i^2 - \frac{2}{N} \overline{x} \sum n_i x_i + \frac{1}{N} \overline{x}^2 . \sum n_i$$

$$= E(X^2) - 2 . E^2(X) + \frac{N}{N} E^2(X)$$

$$= E(X^2) - E^2(X) = V(X)$$

- **L'écart type :** $\sigma(X) = \sqrt{V(X)}$

Scanned with CamScanner
Scanned with CamScanner

Chapitre 3 :

. Les Moyennes :

+ Mode : la valeur de la modalité qui correspond à l'effectif maximal.
c'ad le x_i tq $n_i \geq n_j \quad \forall j = 1, \dots, n$ et $j \neq i$

+ La médiane : La valeur de la modalité qui divise la population en deux parties égaux 50% / 50%

- Le quartile Q_1 25% / 75%
- Le quartile Q_2 50% / 50%
- Le quartile Q_3 75% / 25%

Remarque : L'écart-type nous permet de savoir si les valeurs de la modalité sont proche ou non.

- $\sigma(x) \approx 0 \Rightarrow$ Les valeurs prises sont proch
- Sinon $\Rightarrow$ Les valeurs prises sont mal-distribuées.

1°/ <u>Cas d'une variable discrète</u> .

. On reprend l'exemple de la distribution de 100 familles selon le nombre de pièces du logement.

$$\overline{X} = \frac{1}{N} \cdot \sum x_i n_i = \frac{(10 \times 1) + (15 \times 2) + \dots}{100}$$

$\boxed{\overline{X} = 3{,}86}$

Scanned with CamScanner
Scanned with CamScanner

nb de pieces	1	2	3	4	5	6	7	8	Total
Effectifs	1	15	15	30	10	12	5	3	100
Fréquences	0,1	0,15	0,15	0,3	0,1	0,12	0,05	0,03	1

2°/ Cas d'une variable Continue :

On reprend l'exemple de la répartition des salaires de 200 salariés de l'entreprise Atlas.

Valeurs du Salaire	Effectifs n_i	Fréquences Relatives
$3000 \leq x < 6000$	50	0,25
$6000 \leq x < 9000$	90	0,45
$9000 \leq x < 12000$	30	0,15
$12000 \leq x < 15000$	15	0,075
$15000 \leq x < 18000$	10	0,05
$18000 \leq x \leq 21000$	5	0,025
Total	200	1

$$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} n_i x_i = \frac{(50 \times 450) + (90 \times 7500) + (30 \times 13500) + \cdots}{200}$$

$$\bar{x} = 8400 \text{ DH}$$

- La moyenne harmonique : $\frac{1}{H} = \frac{1}{n} \sum_{i=1}^{k} \frac{n_i}{x_i}$

$$= \left(\frac{n_1}{x_1} + \cdots + \frac{n_k}{x_k} \right)$$

- La moyenne géométrique : $G = \sqrt[n]{\prod_{i=1}^{k} x_i^{n_i}}$

ou $\ln(G) = \sum_{i=1}^{k} f_i \ln(x_i)$ $\quad G = \prod_{i=1}^{k} x_i^{f_i}$

Scanned with CamScanner
Scanned with CamScanner

• La moyenne quadratique Q :

$$Q = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{k} n_i (x_i)^2}$$

$$= \sqrt{\frac{1}{n} \cdot (n_1 x_1^2 + \dots + n_k x_k^2)}$$

Remarque : Quelle que soit la série statistique Considereé on a : $H \leqslant G \leqslant \bar{x} \leqslant Q$.

1.3. Le mode :

• Le cas d'une variable discrète :

Le mode est la modalité de la variable x qui correspond à l'effectif ou la fréquence le plus elevé.

• Remarque :

- Lorsque la distribution a un seul mode : la distribution est unimodale.
- Lorsqu'elle a deux modes : bimodale.
- Lorsqu'elle a deux modes consécutifs : intervalle modal.

On reprend l'exemple de la distribution de 100 familles selon le nombre

le mode $m_0 = x_4 = 4$.

• La valeur $m_0 = e_{i0-1} + (e_{i0} - e_{i0-1}) \left(\frac{d_1}{d_1 + d_2}\right)$ est la valeur modale.

où e_{i0-1} est l'extremité inférieure de la classe modale, et $e_{i0} - e_{i0-1}$ est l'amplitude de la classe modale.

Scanned with CamScanner
Scanned with CamScanner

On reprend l'exemple de la répartition des salaires de 200 salariés de l'entreprise Atlas.

D'après l'histogramme, c'est la classe $[6000, 9000[$ qui a la plus grande ordonnée $h_2 = f_2 = 0,45$

d'où la classe modale est $[e_{i0-1}, e_{i0}[= [6000, 9000[$

est la valeur modale est :

$$m_o = e_{i0-1} + (e_{i0} - e_{i0-1}) \left(\frac{d_1}{d_1 + d_2} \right) = 6000 + (9000 - 6000) \times \left(\frac{0,45 - 0,25}{0,45 - 0,25 + 0,45 - 0,15} \right)$$

$$\boxed{m_o = 7200 \text{ DH}}$$

1.4. La médiane et les quartiles

1/ La médiane : la valeur de la variable statistique telle qu'il ait autant d'observations supérieures et d'observations inférieures et d'observations inférieures à cette valeur, elle partage la série statistique en deux parties égales effectifs.

- Si la variable est discrète :
  - si n est impaire $\Rightarrow$ la valeure médiane est $m_e = x_{p+1}$
  - si n est pair $\Rightarrow$ $m_e = \frac{(x_p + x_{p+1})}{2}$.

- Exemple : afin de réaliser une étude sur la rentabilité sectorielle dans le secteur S_1, une enquête permet d'obtenir les taux de rentabilité en % de 8 entreprises 2,5 ; 3 ; 3,1 ; 4,1 ; 5,4 ; 4,5 ; 5,2 ; 2,5

Scanned with CamScanner
Scanned with CamScanner

- Si la variable est Continue :

La classe médiane est la classe $[e_{i-1}, e_i[$ dont la fréquence cumulée croissante approche 0,5 par excès, càd $F_i > 0,5$

par interprétation linéaire, on a :

$$m_e = e_{i-1} + (e_i - e_{i-1}) \left(\frac{0,5 - F_{i-1}}{F_i - F_{i-1}} \right)$$

$$= e_{i-1} + (e_i - e_{i-1}) \left(\frac{0,5 - F_{i-1}}{f_i} \right)$$

- Exercice (Semblable à l'examen).

on reprend l'exemple de la répartition des salaires.

valeur du salaire	Effectifs n_i	Fréquence relative f_i
	50	0,25
	90	0,45
	30	0,15
	15	0,075
	10	0,05
	5	0,025
Total	200	1

- Remplissage des tableau :

$$f_i = \frac{n_i}{N}$$

$$f_i' = f_i \times \frac{a_r}{a_i}$$

a_i l'amplitude de la classe $a_{i+1} - a_i$

a_r l'amplitude la plus petite

Scanned with CamScanner
Scanned with CamScanner

	n_i	x_i	$n_i x_i$	x_i^2	$n_i x_i^2$	f_i	f_i'	F_i
[3K, 4K[	50	3500	175000	12M250K	612M500K	25%		
[4K, 7K[	90	5500	495000	30M250K	2Md 722M 500K	45%		
[7K, 12K[	30	9500	285000	90M250K	2Md 707M 500K	15%		
[12K, 14K[	15	13000	195000	169M	2Md 35M	7,5%		
[14K, 16K[	10	15000	150000	225M	2Md 250M	5%		
[16K, 21K[	5	18500	92500	342M250K	1Md 711M 250K	2,5%		
Total:	200	65000	13M	4Md & 225M	845Md	100%		

Scanned with CamScanner
Scanned with CamScanner

Partie 2 : Calcul des Probabilités

Chapitre 1 : Analyse Combinatoire

1. Permutation de n objets :

Déf: Soit E un ensemble de n objets, toute suite ordonnée de ces n éléments s'appelle permutation de n objets.

Exemple: les permutations de 3 chiffres 1, 2, 3 sont 123, 132, 213, 231, 312, 321.

on note P_n le nombre de permutations de n objets

Pour $n = 1$, $E = \{a\}$ on peut former qu'une seule permutation $\Rightarrow P_1 = 1$

Pour $n = 2$: $E = \{a, b\}$, on a ab, ba $\Rightarrow P_2 = 2 \times 1 = 2$.

Pour $n = 3$: $E = \{a, b, c\}$, on a abc, acb, bac, bca, cab, cba.

$\Rightarrow$ Donc $P_3 = 6 = 2 \times 3 = P_2 \times 3$

Alors $P_n = n \cdot P_{n-1}$

Pour $n = 4$: on aurra $P_4 = 4 \times P_3 = 4 \times 6 = 24$.

$$\underline{P_n = n \times P_{n-1} = n!}$$

• Supposons une urne contenant les n objets à permuter et une série de n tirages sans remise. le premier objet tiré peut-être l'un quelconque des n objets, le deuxième objet tiré ne peut être que des $n-1$ objets qui restent et ainsi de suite.

Donc $P_n = (n-1) \times (n-2) \times \dots \times 2 \times 1 = n!$

Scanned with CamScanner
Scanned with CamScanner

• Arrangement des K objets :

Déf: Soit E un ensemble de n objets, toute suite ordonnée de K objets prés parmi n ($K \leq n$) s'appelle arrangement de K objets

Exemple : Je m'interesse aux 3 vainqueurs d'une course de 100 m.

• Exemple :

Soit E un ensemble de 4 objets a, b, c, d.
les arrangements de 2 objets ($K=2$) sont :
ab, ba, ac, ca.

On peut aussi former également des arrangements de 3 objets ($K=3$)

$$A_n^K = \frac{n!}{(n-k)!} \quad ; \quad A_n^k = A_{n-1}^k + k.A_{n-1}^{k-1}$$

• Combinaison de K objets :

Déf : Soit E un ensemble de n objets, tout sous ensemble de E composé de K objets prés parmi les n s'appelle combinaisons de K objets.

• Exemple : former des groupes d'etudiants pour des TD
• former des équipes de travail

• Proposition :

$$C_n^k = C_n^{n-k} \quad , \quad C_n^k = C_{n-1}^k + C_{n-1}^{k-1} \quad , \quad C_n^k = \frac{n!}{k!(n-k)!}$$

$$C_n^k = \frac{A_n^k}{k!} \quad , \quad C_n^0 = 1 \; ; \; C_n^1 = n \; ; \; C_n^{n-1} = n.$$

⚠ c'est une expérience dont l'issue est Aléatoire.

Scanned with CamScanner
Scanned with CamScanner

- Exemple : Lancement d'un dé donnant comme résultat.

Soit face 1, face 2, ..., face 6.

2/ lancer une pièce de monnaie trois fois de suite.

- Ensemble Fondamentale :

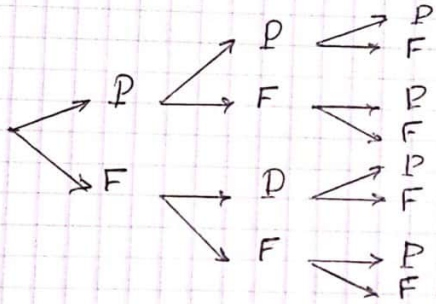
L'ensemble des résultats possibles ou des cas possibles d'une épreuve est appelé Ensemble Fondamental noté Ω.

- Exemple :

E_1 lancement d'une pièce de monnaie $\Omega = \{P, F\}$

E_2 lancement d'une pièce de monnaie trois fois

- Representation des cas possibles :



- <u>Evenement aléatoire</u> :

c'est un sous ensemble de Ω

Exemple : On lance un dé et on s'interesse à l'evenement "Obtenir un chiffre pair" désigné par A cet evenement

$A = \{2, 4, 6\}$.

- L'evenement élementaire : c'est une partie de Ω contenant plus d'un élement de Ω.

Scanned with CamScanner
Scanned with CamScanner

Exemple : à l'issue de l'épreuve E_2, on obtient 2 faces.

- l'evenement certain c'est Ω
- L'evenement impossible est $\{\phi\} = \Omega$
- L'evenement complémentaire $\overline{A}$
- L'evenement « réunion de A et B ».

  $\forall (A,B)$ de Ω, on associe un événement dit « A ou B ». qui se réalise si l'un au moins des événements A et B se réalise, on le note $A \cup B$.

- Exemple : Considérons l'épreuve E_2, et désignons par :

  A = l'evénement "Obtenir 2 Faces"

  B = l'evénement "Obtenir 2 piles"

  $A \cup B = \{FFP, FPF, PFF, PFP, PP, PFF\}$.

Remarque : Si A et B sont deux evénements qui ne peuvent pas se réaliser simultanément, on notera l'évènement $(A \cup B)$ par A et B.

→ L'evénement intersection de A et B.

A tout couple d'événement A, B de Ω, on peut associer l'événement dit « A et B » ou intersection de A et B qui se réalise ssi les deux événements, A et B se réalisent simultanément on note $A \cap B$ ou $A.B$

- Exemple : Soit E_4 l'épreuve qui consiste à tirer au hazard un objet d'une urne contenant

Scanned with CamScanner
Scanned with CamScanner

une boule et un cube noir et une boule et un cube rouge.

Soit l'evenement A = « tirer un cube »
et B « tirer un objet de couleur rouge ».

$A \cap B$ est dans ce cas « tirer un cube rouge »

- evenements incompatibles :

si $A \cap B = \emptyset$ dire $A \cap B$ est un evenement impossible càd A et B ne peuvent pas se réaliser simultan. alors A et B sont dits incompatibles.

$A \cup B = B \cup A$, $A \cap B = B \cap A$

$A \cup (B \cup C) = (A \cup B) \cup C$

$A \cap (B \cap C) = (A \cap B) \cap C$

- Définition d'une probabilité :

$$P : \Omega \longrightarrow [0,1]$$
$$\omega \mapsto P(\omega) \text{ avec } \sum_{\omega \in \Omega} P(\omega) = 1$$

$$P(\omega) = \frac{Card(\omega)}{Card(\Omega)}$$

- Exemple : Lancement d'un dé

Ω	①	②	③	④	⑤	⑥
$P(\omega)$	1/6	1/6	1/6	1/6	1/6	1/6

$P(\Omega) = 1$; $P(\bar{A}) = 1 - P(A)$.

$P(\emptyset) = 1$

$$P(A) = \frac{Card(A)}{Card(\Omega)}$$

- Exemple : Lancement d'un dé

$\Omega = \{1, 2, 3, 4, 5, 6\}$

Scanned with CamScanner
Scanned with CamScanner

$P(1) = P(2) = \dots = P(6) = \frac{1}{6}$, Si on s'interesse aux chiffres impaires , et on désigne pour A cet evenement

$A = \{1, 3, 5\}$; $P(A) = \frac{3}{6} = \frac{1}{2}$

$P(A \cup B) = P(A) + P(A) - P(A \cap B)$.

Exercice : Une urne contient une boule et un cube noirs et une boule et un cube rouge , Quelle est la probabilité de tirer de cette urne un cube ou un objet rouges ?

Scanned with CamScanner
Scanned with CamScanner

$P(1) = P(2) = \dots = P(6) = \frac{1}{6}$, Si on s'interesse aux chiffres impaires, et on désigne pour A cet evenement

$A = \{1, 3, 5\}$; $P(A) = \frac{3}{6} = \frac{1}{2}$

$P(A \cup B) = P(A) + P(A) - P(A \cap B)$.

Exercice : Une urne contient une boule et un cube noirs et une boule et un cube rouge, Quelle est la probabilité de tirer de cette urne un cube ou un objet rouges ?

- Probabilité Conditionnelle :

La réalisation d'un évenement A sachant qu'un autre événement a été réalisé est :

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

- propriétés :

→ $P(\overline{A}/B) = 1 - P(A/B)$

→ $P((A \cup B)/C) = P(A/C) + P(B/C)$ Si $A \cap B = \emptyset$

→ $P((A \cup B)/C) = P(A/C) - P((A \cap B)/C)$ Si $A \cap B = \emptyset$

→ $P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P(C/A \cap B)$.

- Indépendance de deux événements :

L'evenement A est indépendant de l'évènement B si $P(A/B) = P(A)$.

$P(A/B) = \frac{P(A \cap B)}{P(B)} = P(A)$ car $P(A \cap B) = P(A) \cdot P(B)$.

- Propriétés : si A et B sont indépendant, alors :

. A et $\overline{B}$ sont indépendants.

Scanned with CamScanner
Scanned with CamScanner

- $\bar{A}$ et B sont indépendants.
- $\bar{A}$ et $\bar{B}$ sont indépendants.

- Exemple : Une assemblée comprend 400 étudiants qui se répartissent comme suit :
  - 170 marocains et 30 étrangers en 1ère année.
  - 130 marocains et 20 étrangers en 2ème année.
  - 40 marocains et 10 étrangers en 3ème année.

Un étudiant est choisi au hasard, sachant que cet étudiant est marocain, quelle est la probabilité qu'il soit en 2ème année.

Solution :

A : l'évènement de choisir un étudiant marocain

$$P(A) = \frac{Card(A)}{Card(\Omega)} = \frac{170 + 130 + 40}{400}$$

$$= \frac{340}{400} = \frac{34}{40} = \frac{17}{20}.$$

B : L'évènement de choisir un étudiant en 2ème année.

$$P(B) = \frac{150}{400} = \frac{15}{40}$$

$A \cap B$: Choisir un étudiant marocain et en 2ème année

$$P(A \cap B) = \frac{130}{400} = \frac{13}{40}$$

$P(B/A)$ = Choisir de choisir un étudiant en 2ème année sachant qu'il est marocain.

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{13}{40} \times \frac{40}{34} = \frac{13}{34}.$$

Scanned with CamScanner
Scanned with CamScanner

• Exemple :

1/ L'evenement contraire d'un évènement donnée :

Dans une classe, on choisit deux élèves au hazard

A = « Les deux eleves sont des filles.

$\overline{A}$ « au moins l'un des deux élèves et un garçon

* Dans un groupe de suisses et de belges, on discute avec une personne.

B « La personne et un homme belges »

$\overline{B}$ « La personne est soit une femme soit un Suisse.

* Au restaurant, Ahmed prend un plat et un dessert

C « Ahmed prend une viande et une glasse.

$\overline{C}$ « Ahmed ne prend pas de viande or ne prend pas de glasses ».

. Exercice 1 :

Dans une assemblée de 250 personnes, on ne remarque que des hommes partout la cravate ou ayant des yeux bleus, il ya 120 hommes qui porte la cravaté, 85 homemes qui ont des yeux bleaas dont 50 portent la cravate. Calculons :

→ $P(C)$ = avec cravaté.

→ $P(B)$ = yeux bleus.

→ $P(B \cap C)$.

→ $P(B \cup C)$

→ $P(B/C)$

→ $P(C/B)$.

Scanned with CamScanner
Scanned with CamScanner

• Exercice 2 : Lors d'un questionnaire, deux questions étaient posées, 65% des personnes ont répondus « Oui » à la 1ère question et 46% ont répondu « Oui » à la 2ème question, et 46% ont répondu « Oui » aux deux quest :

on note « A » : La personne a répondu "oui" à la 1ère quest.

« B » La personne a répondu "oui" à la 2ème quest »

Solution Ex 1 :

Tableau des Données	Cravaté (C)	Sans Cravate ($\bar{C}$)	Total
Yeux Bleus B	50	35	85
Yeux non bleus $\bar{B}$	70	9	165
Total	120	130	250

$$P(C) = \frac{Card(C)}{Card(\Omega)} = \frac{120}{250} = \frac{12}{25}$$

$$P(B \cap C) = \frac{50}{250} = \frac{1}{5}$$

$$P(B \cup C) = P(B) + P(C) - P(B \cap C)$$

$$= \frac{85}{250} + \frac{120}{250} - \frac{50}{250}$$

$$= \frac{155}{250} = \frac{31}{50}$$

$$P(B \cup C) = \frac{Card(B \cup C)}{Card(\Omega)} = \frac{50 + 70 + 35}{250}$$

$$= \frac{155}{250} = \frac{31}{50}$$

$$P(\bar{B} \cap \bar{C}) = P(\overline{B \cup C}) = 1 - P(B \cup C) = 1 - \frac{31}{50} = \frac{19}{50}.$$

Scanned with CamScanner
Scanned with CamScanner

$$P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B)$$
$$= 1 - (P(A) + P(B) - P(A \cap B))$$
$$= 1 - (0{,}65 + 0{,}51 - 0{,}64) = 0{,}3.$$

- Variable Aléatoire :

Une variable aléatoire X est une variable associée à une expérience aléatoire servant à caractériser les résultats de cette expérience aléatoire.

- variable aléatoire discrète :

Exemple :

Le jet d'un dé, alors X peut prendre de valeurs $\{1, 2, 3, 4, 5, 6\}$

- Loi de probabilité : c'est la relation entre x et $P(X = x)$

x	1	2	3	4	5	6
(x)	1/6	1/6	1/6	1/6	1/6	1/6

$$\sum P(x) = 1$$

- L'espérence mathématique est notée par $E(x) = \sum x\, p(x)$
- La variance est définie par : $V(x) = E(x^2) - E(x)^2$
- L'écart type est : $\sigma(x) = \sqrt{V(x)}$

- La Fonction de répartition est notée par

$$F(x) = P(X \leq x) = \sum^{x} p(x)$$
$$P(a \leq X \leq b) = F(b) - F(a).$$

	1	2	3	4	5	6
$P(x)$	1/6	1/6	1/6	1/6	1/6	1/6
$F(x)$	1/6	2/6	3/6	4/6	5/6	1

$$E(g(x)) = \sum g(x)\, P(x)$$

$$E(x^2) = \sum x^2 p(x).$$

Scanned with CamScanner
Scanned with CamScanner

• Loi de Bernouilli :

Loi de Bernouilli interviennent dans le cas d'une seule expérience aléatoire a laquelle on associe un evenement aléatoire quelconque...
la réalisation de l'evenement au cours de cette expérience est appelée succès et la probabilité de réalisation du succès noté p et non succès noté $q = 1-p$

• Loi Binomiale :

La loi binomiale intervient dans le cas de plusieurs expérience aléatoire identiques et indépendants aux quelles on associe un evenement aléatoire qlq.
et la probabilité de succès p et non succès noté $q = 1-p$.

La variable x qui caractérise le nombre de succès au cours de n expériences aléatoires.
elle prend les valeurs de 0 à n.

$$P(X=x) = C_n^x \, p^x \, q^{n-x}$$

Exemple "Jet d'un dé de 10 faces"

n : nb d'expérience aléatoires
p : La probabilité de succès $p = cte$.

$$\begin{cases} E(x) = np \\ \sigma(x) = \sqrt{npq} \end{cases}$$

Scanned with CamScanner
Scanned with CamScanner

Introduction générale

Le mot statistique tire son origine du latin *statisticus* relatif à l'état (*status*). Il est apparu vers le milieu du XVIIème siècle.

Au pluriel, les « statistiques » signifient un ensemble de données numériques relatives à un groupe d'individus. On parle des statistiques du chiffre d'affaires, de celles du produit intérieur brut, de celles du cours d'un indice boursier, etc.

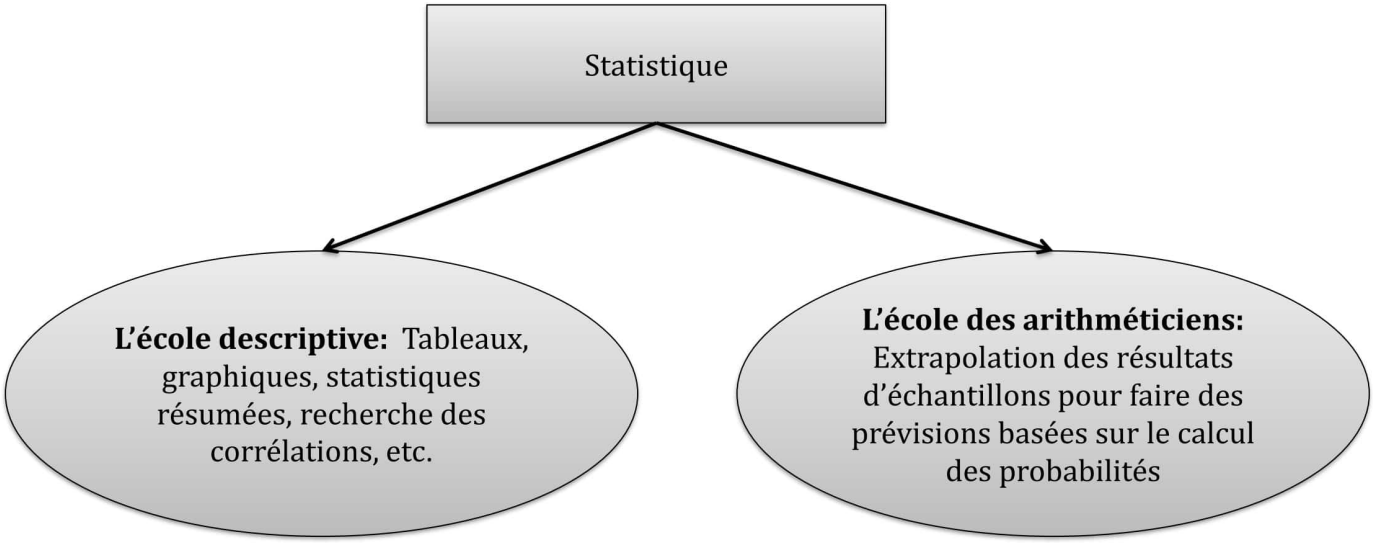
Au singulier, la « statistique » signifie un ensemble des méthodes qui permettent de rassembler, de présenter et d'analyser un ensemble de données numériques.

Scanned with CamScanner

Dans un premier temps, la statistique a été employée dans un sens purement descriptif de recueil ou de collection de faits chiffrés, les *statistiques*.

Dans un second temps, elle a été utilisée pour étendre les résultats et dégager des lois (l'inférence). Elle vise à dégager, à partir de données observées sur quelques individus d'une population, des résultats valables pour l'ensemble de la population. Elle est alors un moyen scientifique d'analyse et de compréhension d'un phénomène étudié. Elle s'applique à l'économie ainsi qu'à toutes les sciences sociales et de la nature.

Scanned with CamScanner



Scanned with CamScanner

La méthodologie statistique

la statistique descriptive consiste à remplacer des données nombreuses par des indicateurs les plus pertinents possibles ainsi qu'à les résumer sous forme de tableaux ou de graphiques.

l'inférence statistique vise, à partir de la description statistique, la mise en évidence de certaines permanences ou lois statistiques qui peuvent constituer des outils de prévision.

la théorie des probabilités (objet du cours du calcul des probabilités en 2ème année) signifie l'analyse mathématique des phénomènes dans lesquels le hasard intervient et qui est utilisée pour déterminer les précisions des estimations de certains paramètres (la théorie de l'estimation) ou des tests de certaines hypothèses (la théorie des tests).

Scanned with CamScanner

Chapitre 1	• La démarche statistique et apport aux sciences économiques et gestion
Chapitre 2	• Les tableaux et représentations graphiques
Chapitre 3	• Les caractéristiques de position et de dispersion
Chapitre 4	• Les indices statistiques
Chapitre 5	• Les séries chronologiques

Scanned with CamScanner

Chapitre 1 : démarche statistique et apport aux sciences économiques et gestion

1- La statistique: des données à l'information

1.1- Notion d'information

1.2 – Rôle de la statistique

2- Apports de la statistique au gestionnaire

3-Terminologie et concepts de base

4- Différents types de caractères et de variable statistiques

Scanned with CamScanner

Chapitre 1 : démarche statistique et apport aux sciences économiques et gestion

1- La statistique: des données à l'information

1.1- Notion d'information

Les données ne constituent pas en soi de l'information. Pour qu'elles le deviennent, elles doivent être porteuses de sens pour l'utilisateur.

→ A l'état brut, les données ne constituent que de l'information potentielle.

Scanned with CamScanner

1.2 – Rôle de la statistique

Le rôle de la statistique est de produire l'information pour aider à la prise de décision. Elle le remplit à partir d'un processus constitué de deux étapes de base:

Etape 1: **collecte** des données de base que l'on nomme les statistiques. A ce stade, elles sont en général inassimilables et donc dénuées de sens pour l'utilisateur.

Etape 2: **extraction** du sens contenu dans ces données.

Scanned with CamScanner

1.2.1- Collecte des données

La première étape de la démarche statistique consiste à sélectionner les critères amenant à collecter les données pouvant devenir les plus significatives.

Exemple:

L'entreprise ATLAS, spécialisée dans la fabrication d'aliments pour animaux domestiques, cherche à adapter sa politique tarifaire aux caractéristiques de sa clientèle.

→ La première étape consistera donc à recueillir les données sur les clients. Mais ceux-ci se différencient les uns des autres par de nombreux aspects: adresse, N°tél, Nom du responsable, Taille…..

Scanned with CamScanner

1.2.2- Extraction du sens

La deuxième étape de la démarche statistique consiste à extraire des informations pertinentes d'une liste de nombres difficile à interpréter par une simple lecture.

Exemple :

Après sélection sur les critères d'activité et de volume, l'entreprise ATLAS a recueilli les 300 données suivantes sous formes de couples (activités, volume):
(grossiste,35315),(détaillant,10112),(particulier,315),…….
(détaillant,21230), (particulier, 5315), (grossistes, 25215)

Scanned with CamScanner

2- Apports de la statistique aux économistes et gestionnaires

L'économiste ou le gestionnaire attend de l'outil statistique une aide à deux niveaux: la compréhension du passé; la prévision de l'avenir.

Cela suppose :

- Une présentation lisible des données recueillies .
- L'analyse des liens mathématiques entre les différentes caractéristiques de la population étudiée aboutissant à des modèles mathématiques sur lequel reposent la prévision.

Scanned with CamScanner

3-Terminologie et concepts de base

3.1- les statistiques et la statistique

3.2- population

Une population est l'ensemble des unités statistiques ou individus étudiés par le statisticien.

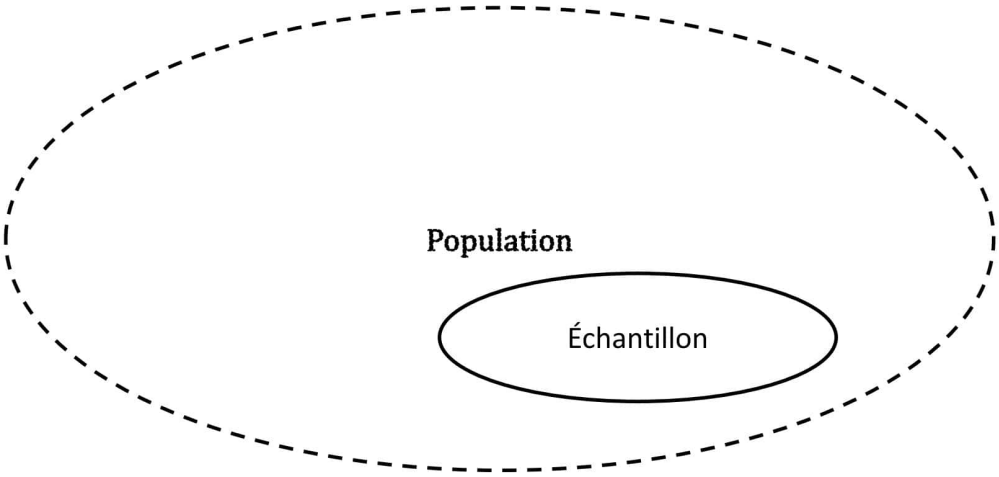
La population étudiée doit être définie avec précision de façon à ce que les enquêteurs ou les différents intermédiaires qui concourent à l'observation des faits, interprètent toujours les instructions de la même façon.

Exemple:
Considérons la population marocaine au 1er janvier 2017 : on devra indiquer si l'on y inclut les étrangers résidant au Maroc, les marocains résidant à l'étranger en précisant le sens du mot résider. Suivant l'objet de l'étude la définition de la population pourra d'ailleurs varier.

Scanned with CamScanner

3.3- échantillon

Il est fréquent que l'on prélève un échantillon dans une population statistique. Le diagramme d'EULER ci-après décrit le lien entre l'échantillon et la population.



Le lien entre l'échantillon et la population

Scanned with CamScanner

3.4- caractères:

Les caractères sont les propriétés étudiées d'un individu (personne, animal, objet ou évènement).

Un caractère est appelé variable statistique, notée X.

Exemple :

Pour décrire la population marocaine on pourra retenir les caractères : sexe, âge, état matrimonial, lieu de résidence, catégorie socioprofessionnelle…s'il s'agit du personnel d'une entreprise, le sexe et l'âge restent intéressants, mais il convient d'y aouter la profession, la qualification, le salaire mensuel, l'ancienneté dans l'entreprise, etc.

Scanned with CamScanner

3.5- modalité:

La modalité est la valeur que peut prendre, ou qu'a réellement pris (on parle aussi d'observation), la variable statistique.

Exemple:

Le caractère « sexe » a deux modalités qui déterminent dans une population le sous-ensemble des individus masculins et le sous-ensemble des individus féminins.

Scanned with CamScanner

4- Différents types de caractères et de variable statistiques

4-1- les caractères qualitatifs

Un caractère qualitatif est un caractère dont les modalités échappent à la mesure. Elles peuvent seulement être constatées

Exemple :

- Sexe de la personne interrogée, situation familiale, numéro de son département de naissance, ...
- L'état du temps constaté à un endroit donné chaque jour (pluvieux, neigeux, beau, venteux, ...)

Scanned with CamScanner

4.2- les caractères quantitatifs

Une variable statistique est quantitative si ses valeurs sont des nombres sur lesquels des opérations arithmétiques telles que somme, moyenne, ... ont un sens.

Exemples :

- Taille, poids, salaire
- Rendement
- Note à un examen
- PNB / habitant, espérance de vie, nombre d'habitants d'un ensemble de pays

Scanned with CamScanner

4.2.1- discret

Une variable statistique est discrète lorsqu'elle ne peut prendre que certaines valeurs isolées dans son intervalle de variation. Il s'agit, en général, de valeurs entières.

Exemples :

- Le nombre de salariés d'une entreprise,
- Le nombre d'enfant à charge d'une famille,
- Le nombre d'accidents du travail survenus dans un établissement,
- Le nombre de ventes d'un certain type d'appareil.

Scanned with CamScanner

*Cas d'une variable discrète

On considère 10 familles d'un quartier avec au moins un enfant de 0 à 17 ans (en âge révolu) : f_1, f_2, …, f_{10}. A chaque famille f_i est associé le nombre d'enfants x_i : $x_1=1$, $x_2=2$, $x_3=1$, $x_4=2$, $x_5=3$, $x_6=4$, $x_7=3$, $x_8=5$, $x_9=1$, $x_{10}=2$.

La population statistique P est l'ensemble des familles du quartier ayant au moins un enfant de 0 à 17 ans (en âge révolu): $P=\{f_1, f_2, \ldots, f_{10}\}$.

Chaque élément f_i de P, c'est-à-dire chacun des 10 familles, est un individu ou une unité statistique.

Scanned with CamScanner

La variable statistique est l'application X : P → {1, 2, 3, …} qui, à chaque famille f_i, associe le nombre $X(f_i)=x_i$ d'enfants.

On constate que X(P)={1, 2, 3, 4, 5} et que les nombres d'observations associées à chacune des valeurs de X(P) sont respectivement : $n_1=3$, $n_2=3$, $n_3=2$, $n_4=1$ et $n_5=1$.

Nombre d'enfants	1	2	3	4	5
Effectifs n_i	3	3	2	1	1

Scanned with CamScanner

4.2.2-continue

Une variable statistique est continue lorsqu'elle peut prendre toutes les valeurs à l'intérieur de son intervalle de variation. Ce nombre de valeurs possibles est toujours infini ; Il est donc nécessaire, avant de classer les observations, de définir les modalités du caractère en groupant en classes les valeurs possibles de la variable statistique.

Exemples de variables continues :

- La taille, le poids, l'Age d'un individu,
- La durée de chômage d'une personne,
- La distance séparant deux points,
- Le débit d'une canalisation.

Scanned with CamScanner

*Cas d'une variable continue

Le quartier industriel Sidi Brahim dispose de 25 PME. A chaque entreprise e_i de la population statistique $P=\{e_1, e_2, e_3, \ldots, e_{25}\}$ est associé le chiffre d'affaires c_i. On définit ainsi une application$X : P \to \mathbb{R}^+$, $e_i \to X(e_i) = c_i$.

Chiffre d'affaires annuel (en millions de DH)	$c < 40$	$40 \leq c < 50$	$50 \leq c < 60$	$60 \leq c < 70$	$70 \leq c < 80$
Effectifs	2	8	8	4	3

Les valeurs possibles de X appartiennent à 5 intervalles : [0, 40[; [40, 50[; [50, 60[; [60, 70[; [70, 80[dont les effectifs respectifs associés sont $n_1=2$; $n_2=8$; $n_3=8$; $n_4=4$; et $n_5=3$.

Scanned with CamScanner

Chapitre 1 : démarche statistique et apport aux sciences économiques et gestion

Exercice 1 :

	Vrai	Faux
La statistique descriptive décrit et analyse		
Une variable statistique discrète ne peut pas être groupée en classes.		
Une variable statistique continue possède des valeurs dans l'ensemble des naturels.		
Une variable statistique est toujours associée à un caractère.		

Scanned with CamScanner

Chapitre 1 : démarche statistique et apport aux sciences économiques et gestion

Exercice 1 :

	Vrai	Faux
La statistique descriptive décrit et analyse		Faux
Une variable statistique discrète ne peut pas être groupée en classes.		Faux
Une variable statistique continue possède des valeurs dans l'ensemble des naturels.	Vrai	
Une variable statistique est toujours associée à un caractère.	Vrai	

Scanned with CamScanner

Chapitre 1 : démarche statistique et apport aux sciences économiques et gestion

Exercice 2 :

Quels sont les individus, la population, les caractères, dans les cas suivants ? Quels autres caractères pourrait-on étudier ?

- Les états des USA et leurs populations respectives.
- Les clients d'une mutuelle d'assurance maladie et leurs professions.
- Les élèves d'un lycée et leurs notes au baccalauréat dans huit matières.

Scanned with CamScanner

Chapitre 1 : démarche statistique et apport aux sciences économiques et gestion

Exercice 2 :

Exemple :

Si on étudie la **production annuelle d'une usine de boîtes d'épinards**, la **population** est **l'ensemble des boîtes produites** durant l'année, et **une boîte** constitue un **individu**. **Un caractère** pourrait être **la quantité de sel présent dans les boîtes**.

Scanned with CamScanner

Chapitre 2: tableaux et représentations graphiques

Scanned with CamScanner

Chapitre 2: tableaux et représentations graphiques

1- classification: vocabulaire

Effectif : l'effectif d'une modalité est le nombre d'observations correspondant à la forme prise par le caractère. On dénombrera par exemple un effectif de 80 particuliers.

Fréquence: correspond à l'effectif de la modalité divisé par l'effectif total. Cela revient à exprimer la part de la modalité dans l'ensemble des observations.

Si le nombre total d'observation est de 300, la fréquence de la modalité « particulier » sera de 80/300= 0,266

Scanned with CamScanner

Chapitre 2: tableaux et représentations graphiques

1. classification: vocabulaire

La distribution est la répartition des effectifs entre les différentes modalités du caractère. Le tableau des (xi;ni) où xi est la valeur du caractère et ni le nombre d'observations de cette valeur est appelé **tableau de distribution**.

Scanned with CamScanner

Chapitre 2: tableaux et représentations graphiques

1. classification: vocabulaire

Les 30 (effectifs) étudiants d'une classe(population) sont ainsi répartis suivant leur année de naissance (caractère ou variable discrète):

Année de naissance	Effectifs
n-1	5
n	10
n+1	8
n+2	7

Distribution

Distribution

Modalités

Fréquences

Scanned with CamScanner

Chapitre 2: tableaux et représentations graphiques

2. **Variables statistiques discrètes**

Soit X une variable discrète qui prend les valeurs numériques $x_1, x_2, \ldots, x_K$ (où $x_1 < x_2 < \ldots < x_K$) et soit n_i les effectifs associés à chaque valeur x_i. Le nombre total d'observations est $n=n_1+n_2+\ldots+n_K$.

Scanned with CamScanner

2. Variables statistiques discrètes

2.1- Présentation en tableau

Tableau des effectifs et des fréquences

Valeurs de X ou modalités	Effectifs	Fréquences relatives
x_1	n_1	$f_1=n_1/n$
x_2	n_2	$f_2=n_2/n$
.	.	.
.	.	.
.	.	.
x_i	n_i	$f_i=n_i/n$
.	.	.
.	.	.
.	.	.
x_{k-1}	n_{k-1}	$f_{k-1}=n_{k-1}/n$
x_k	n_k	$f_k=n_k/n$
	$\sum_{i=1}^{k} n_i = n$	$\sum_{i=1}^{k} f_i = 1$

Scanned with CamScanner

Exemple:

Les données concernant l'activité pourront par exemple être présentées ainsi (en colonne):

Valeurs du caractère	**Nombre d'observations (effectif)**	**Fréquence (relative)**
Particuliers Détaillants Grossistes	80 170 50	0,2666… 0.5666… 0,1666…
Total	N= 300	1

Si l'on appelle X la variable ou le caractère, on pourra nommer:

Xi=[X1, X2, X3] les modalités que peut prendre la variable: X1 pour particulier, X2 pour détaillant et X3 pour grossistes.

ni=[n1,n2,n3] le nombre d'observations pour ces modalités.

N le nombre total d'observations N = $\sum_{i=1}^{n} ni$ soit ici N=n1+n2+n3=300

Scanned with CamScanner

2. Variables statistiques discrètes

2.2- Représentations graphiques

a- Diagramme en bâtons

Dans un système d'axe cartésien, portant en abscisses les valeurs xi de la variable statistique X et en ordonnées les effectifs ni ou les fréquences fi, on trace, à partir de chaque valeur xi, un segment de droite vertical (le bâton) dont la hauteur (la longueur du bâton) est proportionnelle à l'effectif ni ou à la fréquence fi relative et on obtient ainsi le digramme en bâtons.

Scanned with CamScanner

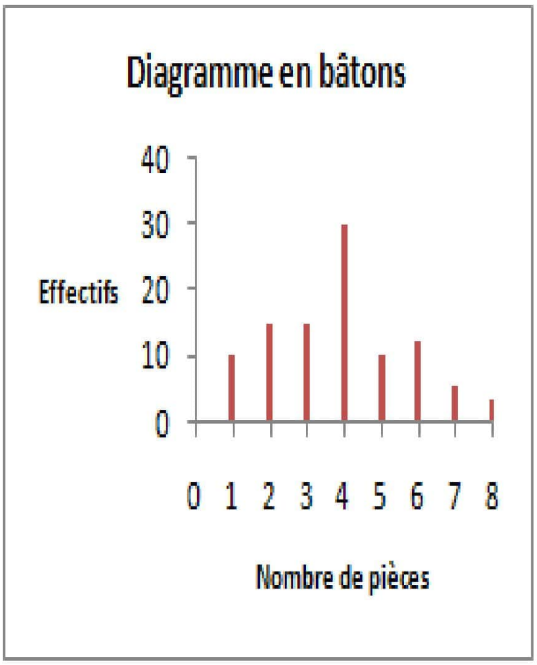
Exemple :

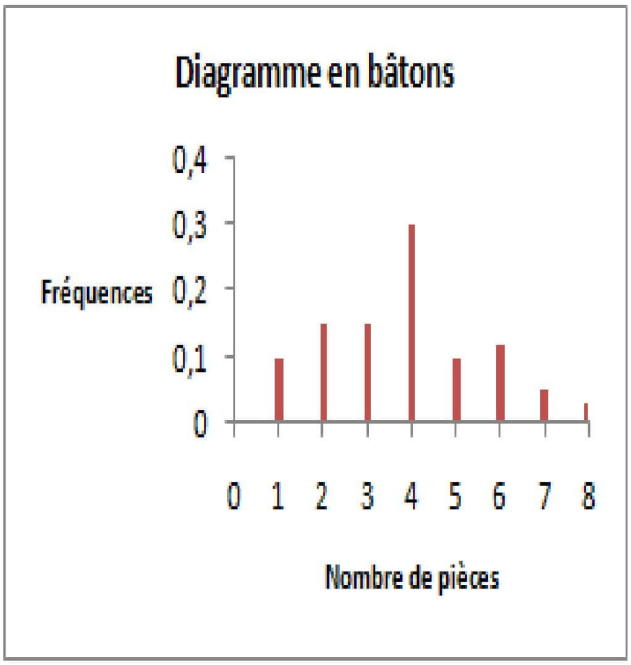
Le tableau ci-dessous présente la distribution de 100 familles selon le nombre de pièces du logement.

Nombre de pièces	1	2	3	4	5	6	7	8	Total
Effectifs	10	15	15	30	10	12	5	3	100
Fréquences	0,1	0,15	0,15	0,3	0,1	0,12	0,05	0,03	1

Scanned with CamScanner

Diagramme en bâtons :



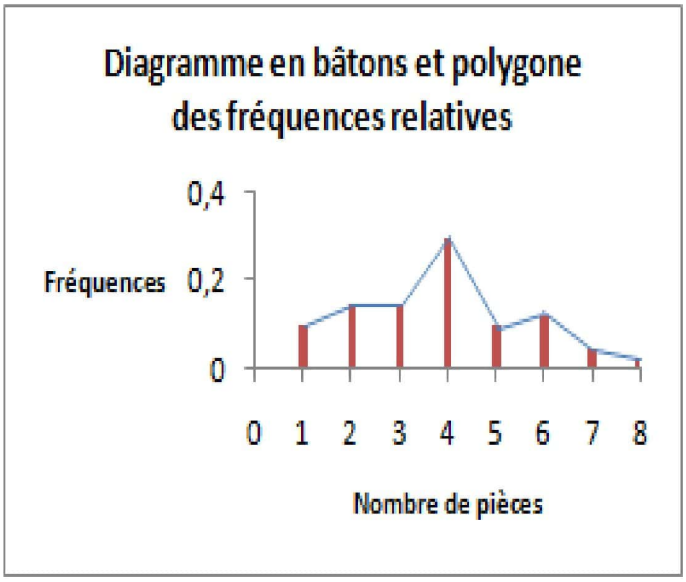


Scanned with CamScanner

2. Variables statistiques discrètes
2.2- Représentations graphiques

b- Le polygone des fréquences

Il est obtenu en joignant les sommets successifs du digramme en bâtons.



Scanned with CamScanner

2. Variables statistiques discrètes

2.3-Effectif cumulé et fréquence relative cumulée

Les tableaux de fréquences (ou d'effectifs) peuvent être modifiés de façon à présenter un résumé des données sous une forme différente.

a- Effectif cumulé (croissant)

L'effectif cumulé croissant jusqu'à une valeur x_h de la variable X est le nombre d'individus pour lesquels la variable X prend une valeur inférieure ou égale à x_h.

b- Fréquence relative cumulée (croissante)

La fraction F_h du nombre total d'individus pour lesquels la variable X prend une valeur inférieure ou égale à x_h est la fréquence relative cumulée croissante des valeurs de X jusqu'à x_h.

Scanned with CamScanner

Valeurs de X u modalités x_i	**Effectifs n_i**	**Effectifs cumulés croissants N_i**	**Fréquences relatives f_i**	**Fréquences relatives cumulées croissantes F_i**
x_1	n_1	$N_1 = n_1$	$f_1 = n_1 / n$	$F_1 = f_1$
x_2	n_2	$N_2 = n_1 + n_2$	$f_2 = n_2 / n$	$F_2 = f_1 + f_2$
x_3	n_3	$N_3 = n_1 + n_2 + n_3$	$f_3 = n_3 / n$	$F_3 = f_1 + f_2 + f_3$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_h	n_h	$N_h = n_1 + n_2 + \ldots + n_h$	$f_h = n_h / n$	$F_h = f_1 + f_2 + \ldots + f_h$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_k	n_k	$N_k = \sum_{i=1}^{k} n_i = n$	$f_k = n_k / n$	$F_k = 1$

Scanned with CamScanner

Exemple :

On reprend le cas de la distribution de 100 familles selon le nombre de pièces du logement.

Nombre de pièces	1	2	3	4	5	6	7	8	Total
Effectifs	10	15	15	30	10	12	5	3	100
Fréquences	0,1	0,15	0,15	0,3	0,1	0,12	0,05	0,03	1

Scanned with CamScanner

Nombre de pièces	Effectifs n_i	Effectifs cumulés croissants N_i	Fréquences relatives f_i	Fréquences relatives cumulées croissantes F_i
1	10	10	0,1	0,1
2	15	25	0,15	0,25
3	15	40	0,15	0,4
4	30	70	0,3	0,7
5	10	80	0,1	0,8
6	12	92	0,12	0,92
7	5	97	0,05	0,97
8	3	100	0,03	1
Total	100	-	1	-

Scanned with CamScanner

3. Variables statistiques continues

Lorsque n est grand (une infinité des valeurs observables d'une variable continue)

l'analyse ci-dessus de cette série d'observations devient fastidieuse (par exemple, la généralisation du diagramme en bâtons)

Subdiviser le domaine des valeurs numériques possibles de la variable statistique continue en K classes consécutives d'amplitudes égales ou non : $[e_0,e_1[$, $[e_1, e_2[$, …, $[e_{k-1}, e_k[$ et de grouper toutes les observations qui appartiennent à une même classe.

Scanned with CamScanner

3. **Variables statistiques continues**

L'intervalle $[e_{i-1}, e_i[$ fermé à gauche, ouvert à droite, est appelé $i^{ème}$ classe ($i = 1, 2, \ldots, k$). Son amplitude est égale à : $a_i = e_i - e_{i-1}$.

Scanned with CamScanner

3. Variables statistiques continues

3.1- Présentation en tableau

Tableau- Effectifs et fréquences relatives

Valeurs de X	Effectifs n_i	Fréquences relatives f_i
$e_0 \leq x < e_1$	n_1	$f_1 = n_1 / n$
$e_1 \leq x < e_2$	n_2	$f_2 = n_2 / n$
.	.	.
.	.	
.	.	.
$e_{i-1} \leq x < e_i$	n_i	$f_i = n_i / n$
.	.	.
.	.	.
.	.	.
$e_{k-1} \leq x < e_k$	n_k	$f_k = n_k / n$

n_i est l'effectif de la classe $[e_{i-1}, e_i[$.
La fréquence relative de la $i^{ème}$ classe est $f_i = n_i/n$.

Scanned with CamScanner

Exemple :

Le tableau ci-dessous présente la répartition des salaires de 200 salariés de l'entreprise Atlas.

Tableau- Effectifs et fréquences relatives

Valeurs du salaire en DH	**Effectifs n_i**	**Fréquences relatives f_i**
$3000 \leq x < 6000$	50	0,25
$6000 \leq x < 9000$	90	0,45
$9000 \leq x < 12000$	30	0,15
$12000 \leq x < 15000$	15	0,075
$15000 \leq x < 18000$	10	0,05
$18000 \leq x < 21000$	5	0,025

Scanned with CamScanner

3. Variables statistiques continues

3.2- Représentations graphiques

a- Histogramme

On peut représenter dans un repère orthogonal la fréquence relative de chaque classe par la surface d'un rectangle qui lui est égale ou proportionnelle.

Cette représentation, nommée histogramme, est ainsi obtenue par la juxtaposition de rectangles dont les bases représentent les différentes classes et dont les surfaces sont proportionnelles aux fréquences relatives des classes.

Scanned with CamScanner

En effet, à la $i^{ème}$ classe, correspond à un rectangle dont la base est l'intervalle $[e_{i-1}, e_i[$ et dont la surface est proportionnelle à la fréquence f_i.

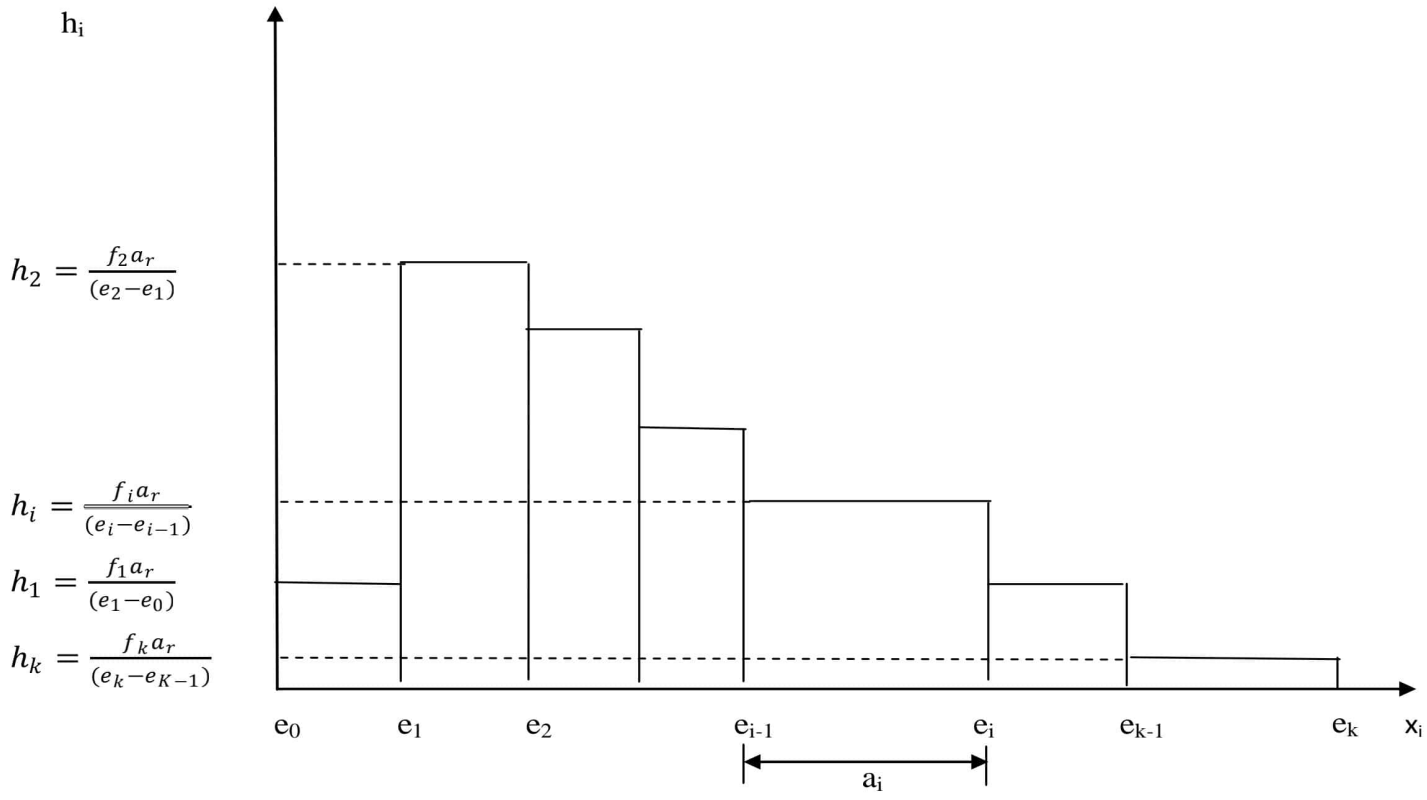
Scanned with CamScanner

les classes ont toutes la même amplitude	les classes sont d'amplitudes inégales
• les hauteurs des rectangles sont égales aux fréquences	• la hauteur du rectangle correspondant à la $i^{ème}$ classe d'amplitude a_i sera $h_i = f_i \cdot a_r / a_i$ où a_r représente l'amplitude de référence (la plus petite des amplitudes).

$f_i' = h_i$ est la fréquence corrigée de la $i^{ème}$ classe.

Scanned with CamScanner

Histogramme des fréquences

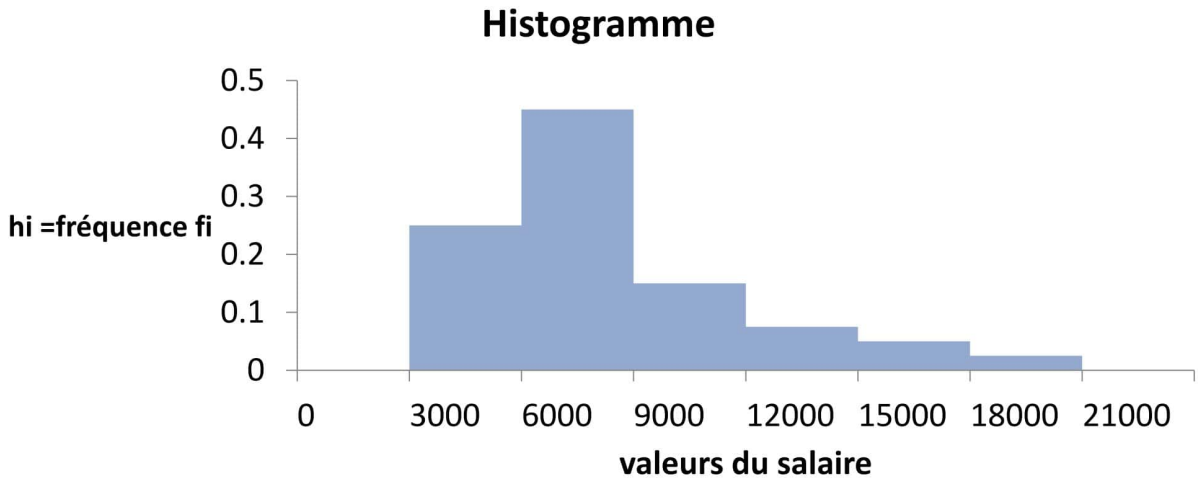


Scanned with CamScanner

Exemples :

cas de classes avec amplitudes égales :

On reprend l'exemple de la distribution des salaires des salariés de l'entreprise Atlas.



Scanned with CamScanner

3. Variables statistiques continues

3.2- Représentations graphiques

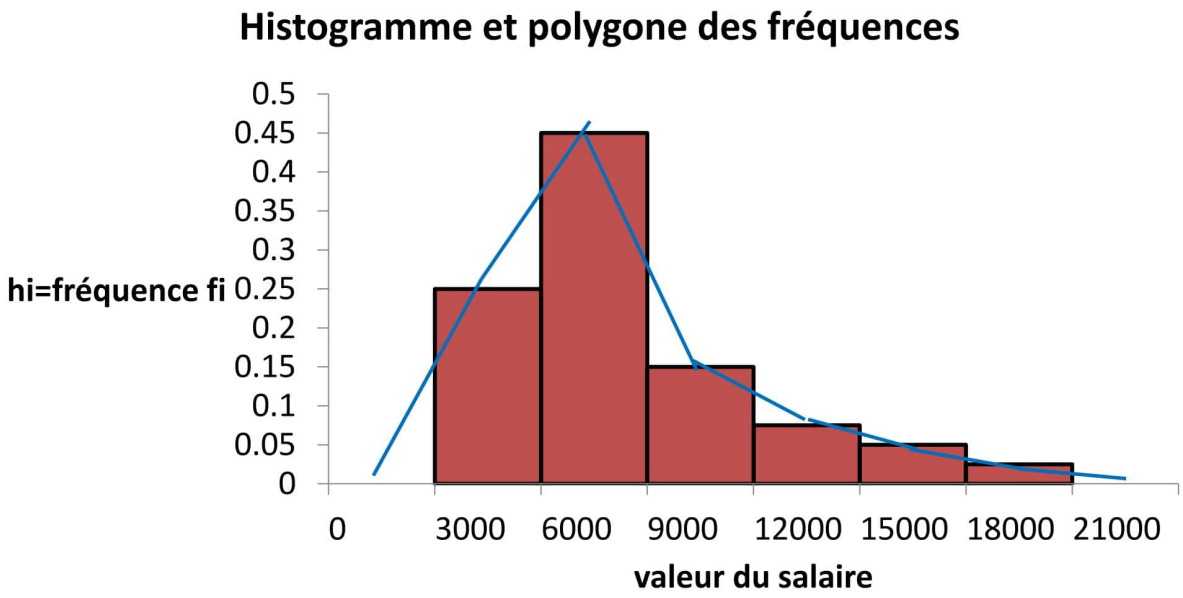
b- Polygone des fréquences :

*Lorsque les classes $[e_0, e_1[$, $[e_1, e_2[$, ..., $[e_{k-1}, e_k[$ ont une même amplitude a ($a = e_i - e_{i-1}$), on joint par des segments de droites les milieux des sommets des rectangles de l'histogramme des fréquences. Le milieu M_i du sommet du rectangle relatif à la classe $[e_{i-1}, e_i[$ a pour abscisse $x_i = (e_{i-1} + e_i) / 2$ et pour ordonnée $y_i = f_i$. Aux extrémités, on ajoute les segments $[M_0 M_1]$ et $[M_k M_{k+1}]$ où M_0 et M_{k+1} ont respectivement pour coordonnées $(e_0 - a / 2, 0)$ et $(e_k + a / 2, 0)$.

Scanned with CamScanner

Exemple 1:

On reprend l'exemple de la distribution des salaires des salariés de l'entreprise Atlas.



Scanned with CamScanner

*Lorsque les classes n'ont pas la même amplitude, on considère parmi les k classes celle ou celles qui ont la plus petite longueur θ. On subdivise l'histogramme des fréquences en sous rectangles de même base égale à la longueur θ. On joint par des segments les milieux des intervalles de longueur θ des sommets des sous rectangles de l'histogramme des fréquences à partir du point $M_0(e_0 - \theta/2, 0)$ jusqu'au point $M_k(e_k + \theta/2, 0)$.

Remarque :

L'aire située sous le polygone doit être égale à l'aire de l'ensemble des rectangles de l'histogramme.

Scanned with CamScanner

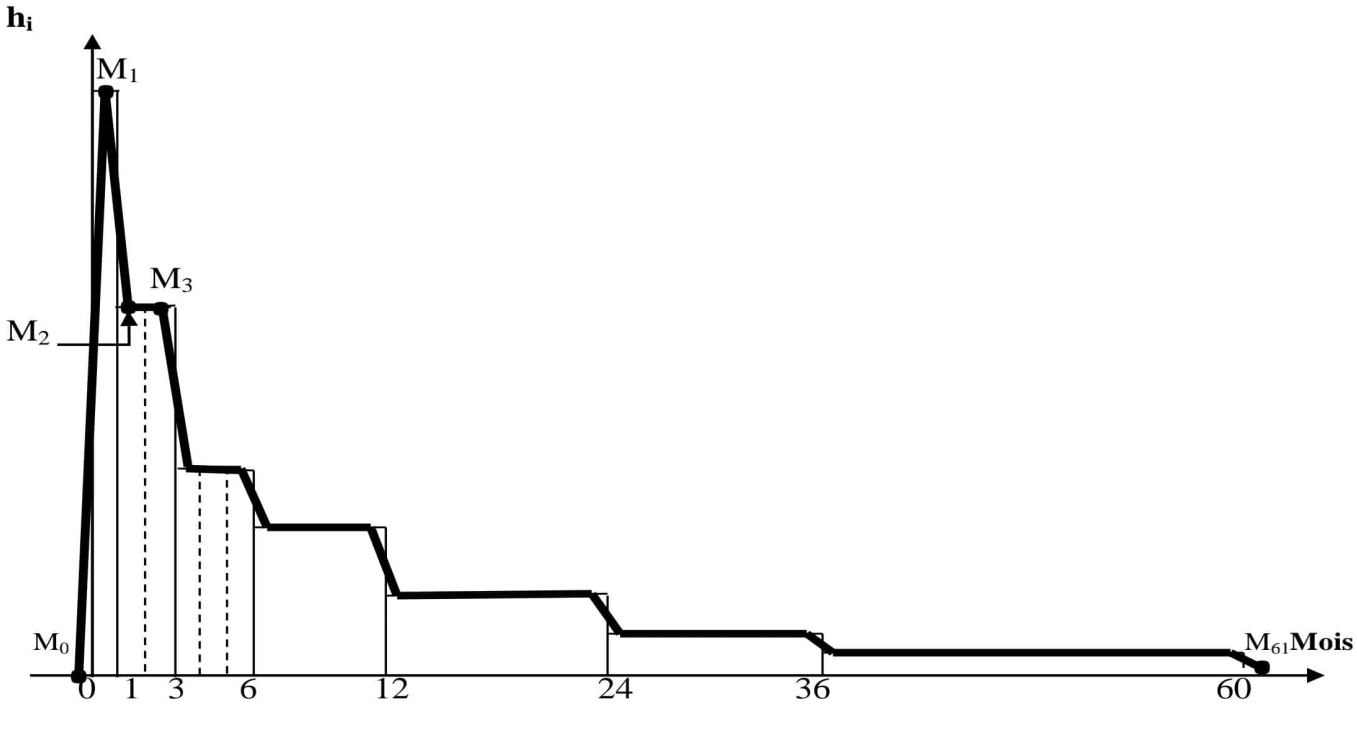
Exemple 2 :

On reprend l'exemple de la distribution de chômeurs inscrits à l'ANAPEC selon l'ancienneté de chômage en Janvier 2015.

Puisque les amplitudes des classes sont inégales et la plus petite amplitude est $\theta=1$, on subdivise alors l'histogramme des fréquences en sous rectangles de même base égale à la longueur $\theta=1$ et on joint par des segments les milieux des intervalles de longueur $\theta=1$ des sommets des sous rectangles de l'histogramme des fréquences à partir du point $M_0(-1/2, 0)$ jusqu'au point M_{61} (60,5 , 0).

Scanned with CamScanner

Histogramme et polygone des fréquences



Scanned with CamScanner

Remarques :

i) L'aire située sous le polygone doit être égale à l'aire de l'ensemble des rectangles de l'histogramme.

ii) Si l'amplitude la plus petite n'est pas un diviseur commun des autres amplitudes, on choisit une valeur de θ, autre que la plus petite amplitude, qui doit être un diviseur commun de toutes les amplitudes afin de faciliter la subdivision des rectangles de l'histogramme en des sous rectangles de base θ.

Scanned with CamScanner

3.3- Effectif cumulé et fréquence relative cumulée

a- Effectif cumulé croissant

L'effectif cumulé croissant de la $h^{ème}$ classe est le nombre d'individus N_h pour lesquels la variable X prend une valeur inférieur à e_h.

$$N_h = \sum_{i=1}^{h} n_i = n_1 + n_2 + \ldots + n_h.$$

Scanned with CamScanner

3.3- Effectif cumulé et fréquence relative cumulée

b- Fréquence relative cumulée (croissante)

La fréquence cumulée croissante de la $h^{ème}$ classe est la fraction F_h du nombre total d'individus pour lesquels la variable X prend une valeur inférieure à e_h.

$$F_h = \sum_{i=1}^{h} f_i = f_1 + f_2 + \ldots + f_h = N_h / n.$$

Scanned with CamScanner

Tableau- Effectifs, fréquences relatives, effectifs cumulés croissants, fréquences relatives cumulées croissantes

Valeurs de X	Effectifs n_i	Effectifs cumulés croissants N_i	Fréquences relatives f_i	Fréquences relatives cumulées croissantes F_i
$e_0 \leq x < e_1$	n_1	$N_1 = n_1$	$f_1 = n_1/n$	$F_1 = f_1$
$e_1 \leq x < e_2$	n_2	$N_2 = n_1 + n_2$	$f_2 = n_2/n$	$F_2 = f_1 + f_2$
$e_2 \leq x < e_3$	n_3	$N_3 = n_1 + n_2 + n_3$	$f_3 = n_3/n$	$F_3 = f_1 + f_2 + f_3$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
$e_{h-1} \leq x < e_h$	n_h	$N_h = n_1 + n_2 + \ldots + n_h$	$f_h = n_h/n$	$F_h = f_1 + f_2 + \ldots + f_h$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
$e_{k-1} \leq x < e_k$	n_k	$N_k = \sum_{i=1}^{k} n_i = n$	$f_k = n_k/n$	$F_k = 1$

Scanned with CamScanner

Exemple :

On reprend le cas de la distribution des chômeurs inscrits à l'ANAPEC selon l'ancienneté de chômage en Janvier 2015.

Ancienneté d'inscription	**Effectif n_i (en milliers)**	**Fréquences f_i**
Moins d'un mois	362	0,168
D'un mois à moins de 3 mois	409,8	0,19
De trois mois à moins de six mois	326,2	0,151
De six mois à moins d'un an	382,2	0,177
D'un an à moins de deux ans	398,6	0,185
De deux ans à moins de trois ans	150,8	0,07
Trois ans ou plus	125,8	0,058
Total	2155,4	1

La classe « trois ans ou plus » est supposée bornée supérieurement par 5 ans (60 mois).

Scanned with CamScanner

Tableau- Effectifs, fréquences relatives, effectifs cumulés croissants, fréquences relatives croissantes

Ancienneté d'inscription	Effectifs n_i	Effectifs cumulés croissants N_i	Fréquences relatives f_i	Fréquences relatives cumulées croissantes F_i
Moins d'un mois	362	362	0,168	0,168
D'un mois à moins de 3 mois	409,8	771,8	0,19	0,358
De trois mois à moins de six mois	326,2	1098	0,151	0,509
De six mois à moins d'un an	382,2	1480,2	0,177	0,686
D'un an à moins de deux ans	398,6	1878,8	0,185	0,871
De deux ans à moins de trois ans	150,8	2029,6	0,07	0,941
Trois ans ou plus	125,8	2155,4	0,058	1
Total	2155,4	-	1	-

Scanned with CamScanner

4- Variables qualitatives

Soit C une variable qualitative qui prend les modalités c_1, c_2, ...,c_K et soit n_i les effectifs associés à chaque modalité c_i. Le nombre total d'observations est $n=n_1+n_2+...+n_K$.

Scanned with CamScanner

4- Variables qualitatives

4.1 Présentation en tableau

Tableau- Effectifs et fréquences relatives

Modalités du caractère C	Effectifs n_i	Fréquences relatives f_i
c_1	n_1	$f_1=n_1/n$
c_2	n_2	$f_2=n_2/n$
.	.	.
.	.	.
.	.	.
c_i	n_i	$f_i=n_i/n$
.	.	.
.	.	.
.	.	.
c_{k-1}	n_{k-1}	$f_{k-1}=n_{k-1}/n$
c_k	n_k	$f_k=n_k/n$
	$\sum_{i=1}^{k} n_i = n$	$\sum_{i=1}^{k} f_i = 1$

n_i est l'effectif de la modalité c_i.

La fréquence relative de la modalité c_i est $f_i=n_i/n$.

Scanned with CamScanner

4- Variables qualitatives

4.2 Représentations graphiques

a- Diagramme en « tuyaux d'orgue »

Dans un repère orthogonal, portant en abscisses les modalités c_i de la variable qualitative C et en ordonnées les effectifs n_i ou les fréquences f_i, on trace, à partir de chaque modalité c_i, un rectangle (tuyau) qui a pour base cette modalité c_i et comme hauteur l'effectif n_i ou la fréquence f_i. La base de chacun des rectangles ne possède aucune signification numérique puisque la variable est qualitative et les bases des rectangles sont égales.

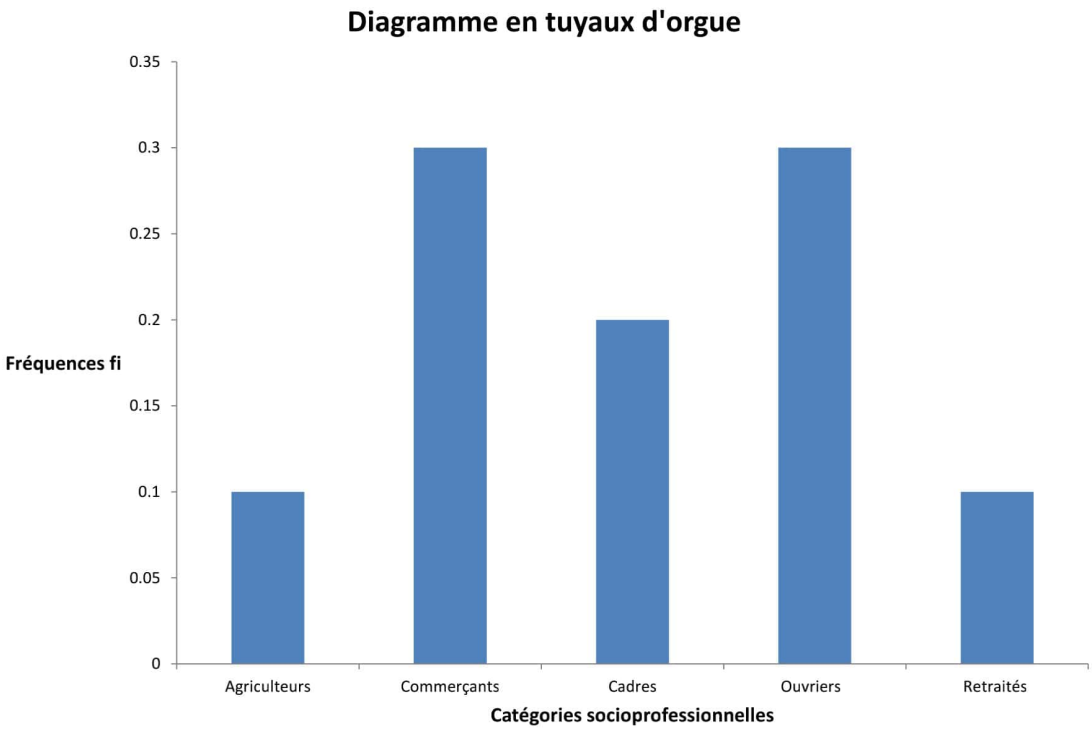
Scanned with CamScanner

Exemple :

Le tableau ci-dessous présente la distribution de 100 personnes selon la catégorie socioprofessionnelle.

Catégories Socioprofessionnelles	Effectifs n_i	Fréquences f_i
Agriculteurs	10	0,1
Commerçants	30	0,3
Cadres	20	0,2
Ouvriers	30	0,3
Retraités	10	0,1
Total	100	1

Scanned with CamScanner



Scanned with CamScanner

4- Variables qualitatives

4.2 Représentations graphiques

b- Diagramme en secteurs

Appelé aussi « Camembert » ou digramme circulaire, le diagramme en secteurs partage un disque en des aires proportionnelles aux effectifs (ou aux fréquences).

Scanned with CamScanner

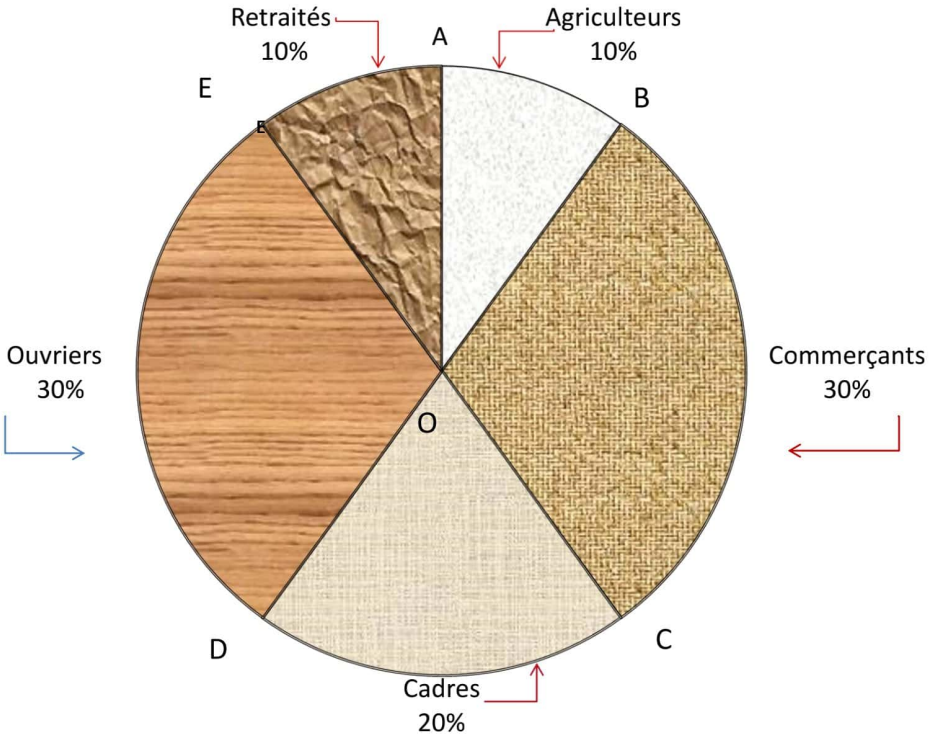
Exemple :

On reprend l'exemple de la distribution de 100 personnes selon la catégorie socioprofessionnelle.

Catégories Socioprofessionnelles	Effectifs n_i	Fréquences f_i
Agriculteurs	10	0,1
Commerçants	30	0,3
Cadres	20	0,2
Ouvriers	30	0,3
Retraités	10	0,1
Total	100	1

Scanned with CamScanner

Diagramme en secteurs



Scanned with CamScanner

Chapitre 3: les caractéristiques de position et de dispersion

- Les indicateurs de tendance centrale ;
- Les indicateurs de dispersion ;
- Les indicateurs de forme.

Scanned with CamScanner

1-Les indicateurs de tendance centrale

Ce sont des indicateurs qui synthétisent l'ensemble d'une série statistique en faisant ressortir une position centrale de la valeur du caractère étudié. Ils regroupent la valeur médiane, différents types de moyennes et la valeur modale.

Scanned with CamScanner

1-1-Moyenne arithmétique $\bar{x}$

La valeur moyenne d'une série statistique est la somme de toutes les observations divisée par le nombre total d'observations. Autrement dit, la valeur moyenne $\bar{x}$ d'une série de n observations $x_1, x_2, \ldots, x_n$ est $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1+x_2+\cdots+x_n}{n}$.

Scanned with CamScanner

i) le cas d'une variable discrète :

Lorsque les n observations d'une variable discrète X prennent l'une des k valeurs x_1, x_2, ..., x_k avec respectivement les effectifs n_1, n_2, ..., n_k, avec $n_1 + n_2 + \ldots + n_k = n$, la valeur moyenne $\bar{x}$ est : $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_k x_k}{n}$

Scanned with CamScanner

On reprend l'exemple de la distribution de 100 familles selon le nombre de pièces du logement.

Nombre de pièces	1	2	3	4	5	6	7	8	Total
Effectifs	10	15	15	30	10	12	5	3	100
Fréquences	0,1	0,15	0,15	0,3	0,1	0,12	0,05	0,03	1

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{8} n_i x_i = \frac{(10x1) + (15x2) + (15x3) + (30x4) + (10x5) + (12x6) + (5x7) + (3x8)}{100} = 3{,}86$$

Scanned with CamScanner

ii) le cas d'une variable continue :

Lorsque les n observations d'une variable continue X sont regroupées en K classes $[e_0, e_1[, [e_1, e_2[, \ldots, [e_{k-1}, e_k[$ d'effectifs respectifs $n_1, n_2, \ldots, n_k$, avec $n_1 + n_2 + \ldots + n_k = n$, la valeur moyenne $\bar{x}$ est : $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_k x_k}{n}$ où x_i est le centre de la $i^{ème}$ classe : $x_i = \frac{e_{i-1} + e_i}{2}$.

Scanned with CamScanner

On reprend l'exemple de la répartition des salaires de 200 salariés de l'entreprise Atlas.

Tableau- Effectifs et fréquences relatives

Valeurs du salaire en DH	Effectifs n_i	Fréquences relatives f_i
$3000 \leq x < 6000$	50	0,25
$6000 \leq x < 9000$	90	0,45
$9000 \leq x < 12000$	30	0,15
$12000 \leq x < 15000$	15	0,075
$15000 \leq x < 18000$	10	0,05
$18000 \leq x < 21000$	5	0,025
Total	200	1

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{6} n_i x_i = \frac{(50x4500) + (90x7500) + (30x10500) + (15x13500) + (10x16500) + (5x19500)}{200} = 8400\,DH$$

Scanned with CamScanner

1-2-Autres moyennes usuelles

i) Moyenne harmonique H :

La moyenne harmonique H de la série de valeurs x_1, x_2, …, x_n, supposées toutes non nulles, est l'inverse de la moyenne arithmétique des inverses des valeurs x_i :

$$\frac{1}{H} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i} = \frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}\right)$$

Scanned with CamScanner

Lorsque les n observations d'une variable X prennent l'une des k valeurs $x_1, x_2, \ldots, x_k$, non nulles, avec respectivement les effectifs $n_1, n_2, \ldots, n_k$, avec $n_1 + n_2 + \ldots + n_k = n$, la valeur moyenne harmonique est :

$$\frac{1}{H} = \frac{1}{n}\sum_{i=1}^{k}\frac{n_i}{x_i} = \frac{1}{n}\left(\frac{n_1}{x_1} + \frac{n_2}{x_2} + \cdots + \frac{n_k}{x_n}\right)$$

$$H = \frac{n}{\sum_{i=1}^{k}\frac{n_i}{x_i}} = \frac{n}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \cdots + \frac{n_k}{x_n}}$$

La définition de la moyenne harmonique présuppose que X ne prend pas la valeur 0.

Scanned with CamScanner

On reprend l'exemple de la distribution de 100 familles selon le nombre de pièces du logement.

$$\frac{1}{H} = \frac{1}{n}\sum_{i=1}^{8}\frac{n_i}{x_i} = \frac{1}{100}\left(\frac{10}{1}+\frac{15}{2}+\frac{15}{3}+\frac{30}{4}+\frac{10}{5}+\frac{12}{6}+\frac{5}{7}+\frac{3}{8}\right) = 0{,}3509$$

$$H = 1/0{,}3509 = 2{,}8498$$

$$\text{Ou encore : } H = \frac{n}{\sum_{i=1}^{8}\frac{n_i}{x_i}} = \frac{100}{\frac{10}{1}+\frac{15}{2}+\frac{15}{3}+\frac{30}{4}+\frac{10}{5}+\frac{12}{6}+\frac{5}{7}+\frac{3}{8}} = 2{,}8498$$

Scanned with CamScanner

ii) Moyenne géométrique G :

La moyenne géométrique G de la série de valeurs x_1, x_2, …, x_n, supposées toutes strictement positives, est définie ainsi :

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 x_2 \dots x_n}$$

Cette formule implique :

$$ln(G) = \frac{1}{n}\sum_{i=1}^{n} \ln(x_i)$$

Scanned with CamScanner

Lorsque les n observations d'une variable X prennent l'une des k valeurs x_1, x_2, …, x_k, strictement positives, avec respectivement les effectifs n_1, n_2, …, n_k, avec $n_1 + n_2 + \ldots + n_k = n$, la valeur moyenne géométrique est :

$$G = \sqrt[n]{\prod_{i=1}^{k} {x_i}^{n_i}} = \sqrt[n]{{x_1}^{n_1} {x_2}^{n_2} \ldots {x_k}^{n_k}}$$

$$G = \prod_{i=1}^{k} {x_i}^{f_i} = {x_1}^{f_1} {x_2}^{f_2} \ldots {x_k}^{f_k}$$

Ces formules impliquent:

$$ln(G) = \sum_{i=1}^{k} f_i ln(x_i)$$

Scanned with CamScanner

On reprend l'exemple de la distribution de 100 familles selon le nombre de pièces du logement.

$$G = \sqrt[n]{\prod_{i=1}^{8} x_i^{n_i}} = \sqrt[100]{1^{10}\ 2^{15}\ 3^{15}\ 4^{30}\ 5^{10}\ 6^{12}\ 7^{5}\ 8^{3}}$$

$$G = \prod_{i=1}^{8} x_i^{f_i} = 1^{0,1}\ 2^{0,15}\ 3^{0,15}\ 4^{0,3}\ 5^{0,1}\ 6^{0,12}\ 7^{0,05}\ 8^{0,03} = 3,3882$$

Scanned with CamScanner

iii) Moyenne quadratique Q :

Lorsque les n observations d'une variable X prennent l'une des k valeurs x_1, x_2,…,x_k avec respectivement les effectifs n_1, n_2, …, n_k, avec $n_1 + n_2 + \ldots + n_k = n$, la valeur moyenne quadratique Q est la racine carrée de la moyenne arithmétique des carrées des valeurs x_i :

$$Q = \sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i (x_i)^2} = \sqrt{\frac{1}{n}(n_1 {x_1}^2 + n_2 {x_2}^2 + \cdots + n_k {x_k}^2)}$$

Scanned with CamScanner

On reprend l'exemple de la distribution de 100 familles selon le nombre de pièces du logement.

$$Q = \sqrt{\frac{1}{n}\sum_{i=1}^{8} n_i (x_i)^2} = \sqrt{\frac{1}{100}[(10 \times 1^2) + (15 \times 2^2) + (15 \times 3^2) + (30 \times 4^2) + (10 \times 5^2) + (12 \times 6^2) + (5 \times 7^2) + (3 \times 8^2)]} = 4{,}2474$$

Scanned with CamScanner

N.B. : Quelle que soit la série statistique considérée où tous les x_i sont strictement positifs :

$$H \leq G \leq \bar{x} \leq Q$$

On reprend l'exemple de la distribution de 100 familles selon le nombre de pièces du logement.

$$H = 2{,}8498 \leq G = 3{,}3882 \leq \bar{x} = 3{,}86 \leq Q = 4{,}2474$$

Scanned with CamScanner

1-3-Le mode

i) le cas d'une variable discrète :

Le mode est la modalité de la variable X qui correspond à l'effectif ou la fréquence le (la) plus élevé(e).

Remarque :

- Lorsque la distribution a un seul mode : distribution est unimodale
- Lorsque la distribution a deux modes: distribution bimodale
- Lorsque la distribution a deux modes consécutifs: intervalle modal

Scanned with CamScanner

On reprend l'exemple de la distribution de 100 familles selon le nombre de pièces du logement.

Avec la valeur x_4=4, l'effectif n_4 = 30 est maximal.

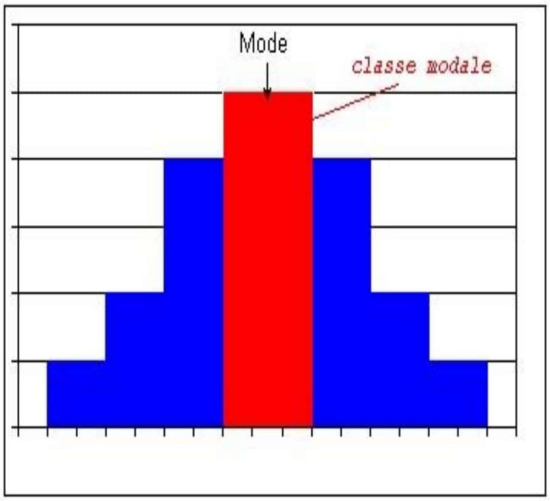
Le mode est alors $m_0 = x_4 = 4$.

Scanned with CamScanner

ii) le cas d'une variable continue :

Lorsque les n observations d'une variable continue X sont regroupées en K classes $[e_0, e_1[, [e_1, e_2[, \ldots, [e_{k-1}, e_k[$ d'effectifs respectifs $n_1, n_2, \ldots, n_k$, avec $n_1 + n_2 + \ldots + n_k = n$, la classe modale est celle qui, dans l'histogramme des fréquences , a la plus grande ordonnée.

L'intervalle $[e_{i0-1}, e_{i0}[$ est la classe modale si et seulement si :



$$h_{i0} = \frac{f_{i0} a_r}{(e_{i0} - e_{i0-1})} > h_i = \frac{f_i a_r}{(e_i - e_{i-1})} \forall\, i \neq io$$

Scanned with CamScanner

La valeur $m_o = e_{io-1} + (e_{io} - e_{io-1})(\frac{d_1}{d_1+d_2})$ est la valeur modale, où : e_{io-1} est l'extrémité inférieure de la classe modale, $e_{io} - e_{io-1}$ est l'amplitude de la classe modale, $d_1 = f_i - f_{i-1}$ et $d_2 = f_i - f_{i+1}$ lorsque les classes ont la même amplitude ou $d_1 = f'_i - f'_{i-1}$ et $d_2 = f'_i - f'_{i+1}$ lorsque les classes ont des amplitudes différentes.

Scanned with CamScanner

On reprend l'exemple de la répartition des salaires de 200 salariés de l'entreprise Atlas.

D'après l'histogramme, c'est la classe [6000 ; 9000[qui a la plus grande ordonnée : $h_2 = f_2 = 0{,}45$.

D'où la classe modale est $[e_{io-1}, e_{io}[= [6000 ; 9000[$ et la valeur modale est :

$$m_o = e_{io-1} + (e_{io} - e_{io-1})\left(\frac{d_1}{d_1 + d_2}\right) = 6000 + (9000 - 6000)\left(\frac{0{,}45 - 0{,}25}{0{,}45 - 0{,}25 + 0{,}45 - 0{,}15}\right) = 7200\ DH$$

Scanned with CamScanner

Quant à l'exemple de la distribution des chômeurs inscrits à l'ANAPEC selon l'ancienneté de chômage en Janvier 2015, c'est la classe [0 ; 1[qui a la plus grande ordonnée :

$$h_1 = f_1 \,.\, a_r / a_1 = f'_1 = 0{,}168.$$

D'où la classe modale est $[e_{io-1}, e_{io}[= [0 ; 1[$ et la valeur modale est :

$$m_o = e_{io-1} + (e_{io} - e_{io-1})\left(\frac{d_1}{d_1+d_2}\right) = 0 + (1-0)\left(\frac{0{,}168-0}{0{,}168-0+0{,}168-0{,}095}\right) = 0{,}6971\ mois.$$

Scanned with CamScanner

1-4- La médiane et les quartiles

a- La médiane

La médiane est la valeur de la variable statistique telle qu'il ait autant d'observations supérieures et d'observations inférieures à cette valeur. Elle partage la série statistique en deux parties d'égal effectif.

Scanned with CamScanner

i) le cas d'une variable discrète :

*Soit une série statistique de n observations x_1, x_2, …, x_n connues, distinctes ou non et classées par ordre de valeurs croissantes : $x_1 \leq x_2 \leq \ldots \leq x_n$.

On appelle médiane la valeur m_e, telle que : (nombre d'observations inférieures ou égales à m_e) ≥ n/2 et (nombre d'observations supérieures ou égales à m_e) ≥ n/2. Autrement dit, il y a au moins 50 % des observations qui ont une valeur inférieure ou égale à m_e la médiane et il y a au moins 50% des observations qui ont une valeur supérieure ou égale à m_e.

Sa définition dépend de la parité de n :

- Si n est un nombre impair (n = 2 p + 1), la valeur médiane sera $m_e = x_{p+1}$;

- Si n est un nombre pair (n = 2 p), la valeur médiane sera $m_e = (x_p + x_{p+1})/2$.

Scanned with CamScanner

Exemple:

Afin de réaliser une étude sur la rentabilité sectorielle dans le secteur S1, une enquête permet d'obtenir les taux de rentabilité en % de 8 entreprises : 2,5 ; 3 ; 3,1 ; 4,1 ; 5,4 ; 4,5 ; 5,2 ; 2,5.

La valeur médiane de cet échantillon est obtenue en classant ces 8 valeurs par ordre croissant :

Rang de l'observation i	1	2	3	4	5	6	7	8
Rentabilité financière (en%) x_i	2,5	2,5	3,0	3,1	4,1	4,5	5,2	5,4

Scanned with CamScanner

Le nombre d'observations n = 8 étant un nombre pair (n = 2 p avec p = 4) la valeur médiane m_e correspond à $(x_p + x_{p+1})/2$soit:

$(x_4 + x_5)/2 = (3{,}1 + 4{,}1)/2 = 3{,}6.$

Scanned with CamScanner

Pour comparer la rentabilité des firmes de ce secteur S1 à celle des firmes du secteur S2, on a prélevé dans ce dernier un échantillon de 9 entreprises : 3 ; 3,5 ; 3,8 ; 4,1 ; 4,1 ; 4,5 ; 5,5 ; 5,5 ; 4,2.

Pour obtenir la valeur médiane de ce deuxième échantillon, on réécrit les valeurs de cette série dans l'ordre croissant :

Rang de l'observation i	1	2	3	4	5	6	7	8	9
Rentabilité financière (en%) x_i	3	3,5	3,8	4,1	4,1	4,2	4,5	5,5	5,5

Scanned with CamScanner

La taille de l'échantillon étant un nombre impair ($n = 2p + 1$ avec $p = 4$), la valeur médiane m_e correspond à x_{p+1} : $m_e = x_5 = 4{,}1$.

Scanned with CamScanner

ii) le cas d'une variable continue :

Lorsque les n observations d'une variable continue X sont regroupées en K classes $[e_0, e_1[$, $[e_1, e_2[$, …, $[e_{k-1}, e_k[$ d'effectifs respectifs n_1, n_2, …, n_k, avec $n_1 + n_2 + \ldots + n_k = n$, la classe médiane est la classe $[e_{i-1}, e_i[$ dont la fréquence cumulée croissante approche 0,5 par excès, c'est-à-dire $F_i > 0{,}5$.

Scanned with CamScanner

Par interpolation linéaire, on a :

$$m_e = e_{i-1} + (e_i - e_{i-1})\left(\frac{0{,}5 - F_{i-1}}{F_i - F_{i-1}}\right)$$

$$= e_{i-1} + (e_i - e_{i-1})(\frac{0{,}5 - F_{i-1}}{f_i})$$

Scanned with CamScanner

On reprend l'exemple de la répartition des salaires de 200 salariés de l'entreprise Atlas.

la classe médiane est la classe [6000, 9000[dont la fréquence cumulée croissante approche 0,5 par excès, c'est-à-dire $F_2 = 0{,}70 > 0{,}5$.

Scanned with CamScanner

D'où : la classe médiane est : [e_1, e_2[=[6000 ; 9000[et, par interpolation linéaire, on a :

$$m_e = e_1 + (e_2 - e_1)\left(\frac{0{,}5 - F_1}{F_2 - F_1}\right) = e_1 + (e_2 - e_1)\left(\frac{0{,}5 - F_1}{f_2}\right)$$

$$m_e = 6000 + (9000 - 6000)\left(\frac{0{,}5 - 0{,}25}{0{,}70 - 0{,}25}\right)$$

$$= 6000 + (9000 - 6000)\left(\frac{0{,}5 - 0{,}25}{0{,}45}\right) = 7666{,}6667$$

Scanned with CamScanner

b- Les quantiles

Le quantile est un indicateur de position. Il est une généralisation de la notion de médiane.

Le quantile d'ordre α ($0 \leq \alpha \leq 1$), noté x_α, est tel qu'une proportion α des individus ait une valeur du caractère X inférieure ou égale à x_α. Le quantile $x_{0,5}$ est égal à la médiane.

Les quantiles les plus courants sont : les quartiles, les déciles et les centiles.

Scanned with CamScanner

Les quartiles sont les valeurs d'une série ou d'une distribution statistique rangées par ordre croissant (ou décroissant) qui partagent l'effectif total en quatre parties égales. Ils sont les quantiles d'ordre ¼, ½ et ¾. Ils sont ainsi notés : Q_1 = premier quartile = $x_{0,25}$;

Q_2 = deuxième quartile = médiane = $x_{0,5}$

et Q_3 = troisième quartile = $x_{0,75}$ ($Q_1 \leq Q_2 \leq Q_3$).

Scanned with CamScanner

i) le cas d'une variable discrète :

Les trois quartiles Q_1, Q_2 et Q_3 sont des nombre qui partagent les n observations en 4 classes qui ont chacune un effectif sensiblement égal à n/4. Le second quartile coïncide évidemment avec la médiane.

Scanned with CamScanner

- *Premier quartile*. Le premier quartile Q_1 est défini comme suit :

  $Q_1 = (x_{n/4} + x_{n/4+1})/2$ si n/4 est un entier ;

  $Q_1 = x_{[n/4]+1}$ (où [n/4] est la partie entière de n/4) si n/4 n'est pas un entier.

On a la propriété : il y a au moins 25 % des observations qui ont une valeur inférieure ou égale à Q_1 et il y a au moins 75% des observations qui ont une valeur supérieure ou égale à Q_1.

Scanned with CamScanner

- *Deuxième quartile*. Sa valeur Q_2 est égale à la médiane m_e.
- *Troisième quartile*. Sa valeur Q_3 est définie comme suit :

  $Q_3 = (x_{3n/4} + x_{3n/4+1})/2$ si $3n/4$ est un entier ;

  $Q_3 = x_{[3n/4]+1}$ (où $[3n/4]$ est la partie entière de $3n/4$) si $3n/4$ n'est pas un entier.

On a la propriété : il y a au moins 75 % des observations qui ont une valeur inférieure ou égale à Q_3 et il y a au moins 25% des observations qui ont une valeur supérieure ou égale à Q_3.

Scanned with CamScanner

Exemple :

Pour l'échantillon des 8 firmes issues de S1, le nombre d'observations n étant égal à 8, n/4 est l'entier 2.

Par suite $Q_1= (x_{n/4} +x_{n/4+1})/2=(x_2 +x_3)/2=(2,5+3)/2=2,75$.

3n/4 étant l'entier 6, on a:

$Q_3=(x_{3n/4} +x_{3n/4+1})/2=(x_6 +x_7)/2=(4,5+5,2)/2=4,85$.

Pour les 9 firmes issues de S2 le nombre d'observations n n'est pas divisible par 4 car n/4=9/4=2,25. Et [n/4] la partie entière de n/4 étant égale à 2, on a $Q_1= x_{[n/4]+1}= Q_1= x_3=3,8$. De même, on a $Q_3= x_{[3n/4]+1}= x_7=4,5$.

Scanned with CamScanner

ii) le cas d'une variable continue :

- *Premier quartile*. La classe qui contient le premier quartile est la classe $[e_{i-1}, e_i[$ dont la fréquence cumulée croissante approche 0,25 par excès, c'est-à-dire $F_i > 0{,}25$.

Par interpolation linéaire on a :

$$Q_1 = e_{i-1} + (e_i - e_{i-1})\left(\frac{0{,}25 - F_{i-1}}{F_i - F_{i-1}}\right)$$

- *Deuxième quartile*. Sa valeur Q_2 est égale à la médiane m_e.
- *Troisième quartile*. La classe qui contient le troisième quartile est la classe $[e_{i-1}, e_i[$ dont la fréquence cumulée croissante approche 0,75 par excès, c'est-à-dire $F_i > 0{,}75$.

Par interpolation linéaire on a :

$$Q_3 = e_{i-1} + (e_i - e_{i-1})\left(\frac{0{,}75 - F_{i-1}}{F_i - F_{i-1}}\right)$$

Scanned with CamScanner

Exemple :

On reprend l'exemple de la répartition des salaires de 200 salariés de l'entreprise Atlas.

- La classe qui contient le premier quartile est la classe [6000, 9000[telle que F(9000) = 0,7 > 0,25 et F(6000) = 0,25 ≤ 0,25. Par interpolation linéaire on a :

$$Q_1 = e_{i-1} + (e_i - e_{i-1})\left(\frac{0{,}25 - F_{i-1}}{F_i - F_{i-1}}\right) = 6000 + (9000 - 6000)\left(\frac{0{,}25 - 0{,}25}{0{,}7 - 0{,}25}\right) = 6000.$$

- *Deuxième quartile*. Sa valeur Q_2 est égale à la médiane $m_e = 7666{,}6667$.
- *Troisième quartile*. Sa valeur estimée Q_3 est définie par $F(Q_3) = 0{,}75$. La classe qui contient le troisième quartile est la classe [9000, 12000[telle que F(12000)=0,85 > 0,75 et F(9000)=0,7 ≤ 0,75. Par interpolation linéaire on a :

$$Q_3 = e_{i-1} + (e_i - e_{i-1})\left(\frac{0{,}75 - F_{i-1}}{F_i - F_{i-1}}\right) = 9000 + (12000 - 9000)\left(\frac{0{,}75 - 0{,}7}{0{,}85 - 0{,}7}\right) \approx 10000.$$

Scanned with CamScanner

2- Les indicateurs de dispersion

2-1-La dispersion en termes d'écarts moyens

Le calcul des écarts moyens se basent sur les caractéristiques de tendance centrale, telles que : la valeur médiane m_e ou la valeur moyenne $\bar{x}$. Il considère ces écarts en valeur absolue ou en puissance paire afin de traduire correctement la dispersion.

Scanned with CamScanner

a-Ecarts absolus moyens

i) Ecart absolu moyen par rapport à la valeur médiane m_e

L'écart absolu moyen par rapport à la valeur médiane m_e d'une série de n observations $x_1, x_2, \ldots, x_n$ est

$$e_{m_e} = \frac{1}{n}\sum_{i=1}^{n}|x_i - m_e| = \frac{1}{n}(|x_1 - m_e| + |x_2 - m_e| + \cdots + |x_n - m_e|)$$

Lorsque les n observations d'une variable X prennent l'une des k valeurs $x_1, x_2, \ldots, x_k$ avec respectivement les effectifs $n_1, n_2, \ldots, n_k$, avec $n_1 + n_2 + \ldots + n_k = n$, l'écart absolu moyen par rapport à la valeur médiane m_e est :

$$e_{m_e} = \frac{1}{n}\sum_{i=1}^{k} n_i|x_i - m_e| = \frac{1}{n}(n_1|x_1 - m_e| + n_2|x_2 - m_e| + \cdots + n_k|x_k - m_e|)$$

Scanned with CamScanner

ii) Ecart absolu moyen par rapport à la moyenne arithmétique $\bar{x}$

L'écart absolu moyen par rapport à la moyenne arithmétique $\bar{x}$ d'une série de n observations $x_1, x_2, \ldots, x_n$ est

$$e_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| = \frac{1}{n}(|x_1 - \bar{x}| + |x_2 - \bar{x}| + \cdots + |x_n - \bar{x}|)$$

Lorsque les n observations d'une variable X prennent l'une des k valeurs $x_1, x_2, \ldots, x_k$ avec respectivement les effectifs $n_1, n_2, \ldots, n_k$, avec $n_1 + n_2 + \ldots + n_k = n$, l'écart absolu moyen par rapport à la moyenne arithmétique $\bar{x}$ est :

$$e_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{k} n_i|x_i - \bar{x}| = \frac{1}{n}(n_1|x_1 - \bar{x}| + n_2|x_2 - \bar{x}| + \cdots + n_k|x_k - \bar{x}|)$$

Scanned with CamScanner

b- Variance et écart-type

Ils mesurent la dispersion des données autour de la moyenne.

La variance d'une variable statistique X est la moyenne arithmétique des carrées des écarts à la moyenne arithmétique :

$$\text{Var(X)} = \text{V(X)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - \bar{x}^2$$

Lorsque les n observations d'une variable X prennent l'une des k valeurs $x_1, x_2, \ldots, x_k$ avec respectivement les effectifs $n_1, n_2, \ldots, n_k$, avec $n_1 + n_2 + \ldots + n_k = n$, la variance de X est :

$$\text{V(X)} = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{k} n_i {x_i}^2 - \bar{x}^2$$

Lorsque X est une variable continue les x_i sont les centres des classes $[e_{i-1} ; e_i[$: $x_i = \frac{e_{i-1}+e_i}{2}$

Scanned with CamScanner

L'écart-type est la racine carrée de la variance : $\sigma_X = \sqrt{V(X)}$

L'écart-type est exprimé dans la même unité que les observations, alors que la variance s'exprime dans le carré de cette unité.

N.B. : L'écart-type donne plus de poids aux observations extrêmes que l'écart absolu moyen à la moyenne : $\sigma_X \geq e_{\bar{x}}$.

Scanned with CamScanner

Exemple 1 (le cas d'une variable discrète):

On reprend l'exemple de la distribution de 100 familles selon le nombre de pièces du logement.

x_i	n_i	$n_i x_i^2$	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$n_i(x_i - \overline{x})^2$
1	10	10	-2,86	8,1796	81,796
2	15	60	-1,86	3,4596	51,894
3	15	135	-0,86	0,7396	11,094
4	30	480	0,14	0,0196	0,588
5	10	250	1,14	1,2996	12,996
6	12	432	2,14	4,5796	54,9552
7	5	245	3,14	9,8596	49,298
8	3	192	4,14	17,1396	51,4188
Total	100	1804	-	-	314,04

Scanned with CamScanner

La variance de X est : $V(X) = \frac{1}{n}\sum_{i=1}^{k} n_i {x_i}^2 - \bar{x}^2 = \left(\frac{1804}{100}\right) - 3{,}86^2 = 3{,}1404\ pièces^2$

Ou encore : $V(X) = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \bar{x})^2 = \frac{314{,}04}{100} = 3{,}1404\ pièces^2$

L'écart-type de X est : $\sigma_X = \sqrt{V(X)} = 1{,}7721\ pièces.$

Scanned with CamScanner

Exemple 2 (le cas d'une variable continue):

On reprend l'exemple de la répartition des salaires de 200 salariés de l'entreprise Atlas.

Valeurs du salaire en milliers de DH	x_i **(le centre de la classe** $[e_{i-1}; e_i[$)	n_i	$n_i x_i$	$n_i x_i^2$	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$n_i(x_i - \overline{x})^2$
$3 \leq x < 6$	4,5	50	225	1012,5	-3,9	15,21	760,5
$6 \leq x < 9$	7,5	90	675	5062,5	-0,9	0,81	72,9
$9 \leq x < 12$	10,5	30	315	3307,5	2,1	4,41	132,3
$12 \leq x < 15$	13,5	15	202,5	2733,75	5,1	26,01	390,15
$15 \leq x < 18$	16,5	10	165	2722,5	8,1	65,61	656,1
$18 \leq x < 21$	19,5	5	97,5	1901,25	11,1	123,21	616,05
Total	-	200	1680	16740	-	-	2628

Scanned with CamScanner

La moyenne arithmétique est $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{1680}{200} = 8{,}4\ milliers\ de\ DH.$

La variance de X est :

$$V(X) = \frac{1}{n}\sum_{i=1}^{k} n_i {x_i}^2 - \bar{x}^2 = \left(\frac{16740}{200}\right) - 8{,}4^2 = 13{,}14\ (milliers\ de\ DH)^2$$

Ou encore : $V(X) = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \bar{x})^2 = \frac{2628}{200} = 13{,}14\ (milliers\ de\ DH)^2$

L'écart-type de X est : $\sigma_X = \sqrt{V(X)} = 3{,}6249\ milliers\ de\ DH.$

Scanned with CamScanner

c- Moment centré d'ordre h :

Le moment centré d'ordre h est la moyenne arithmétique des écarts des valeurs de X à leur moyenne arithmétique élevés à la puissance h :

$$\mu_h = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^h$$

Lorsque les n observations d'une variable X prennent l'une des k valeurs x_1, x_2, …,x_k avec respectivement les effectifs n_1, n_2, …, n_k, avec $n_1 + n_2 + \ldots + n_k = n$, le moment centré d'ordre h est :

$$\mu_h = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^h$$

Remarque : La variance est le moment centré d'ordre 2.

Scanned with CamScanner

2-2- Coefficients de dispersion relative de la distribution

i) coefficient interquartile = $\frac{Q_3 - Q_1}{Q_2}$

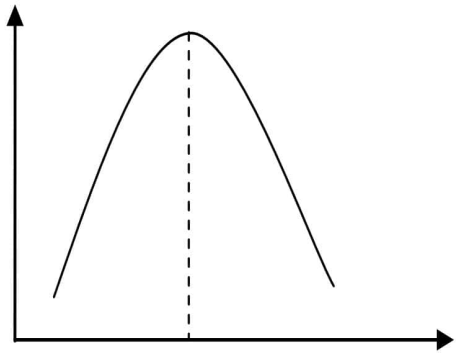
ii) coefficient de variation = $\frac{\sigma_X}{\bar{x}}$. Cet indicateur est utile pour comparer la dispersion d'un caractère sur des populations ayant des valeurs moyennes très différentes ou exprimées dans des unités de mesure différente (monnaies différentes...).

Scanned with CamScanner

3- Indicateurs de forme

3-1- Coefficient d'asymétrie

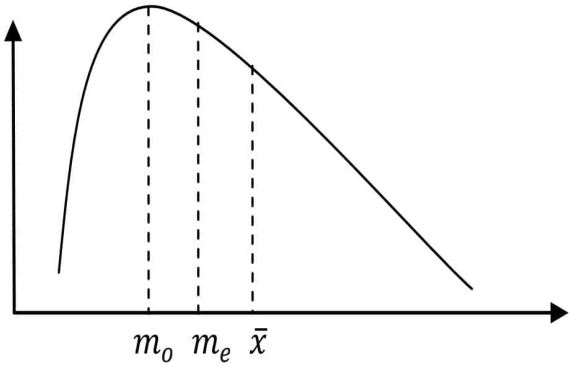
i) Une distribution symétrique



$m_o = m_e = \bar{x}$

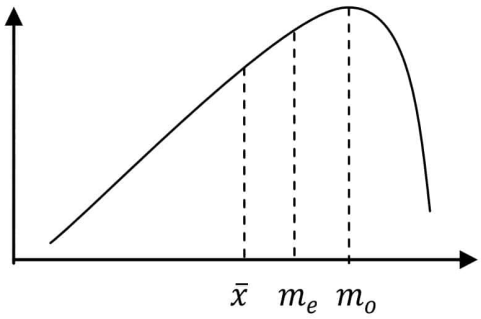
Scanned with CamScanner

ii) Une distribution asymétrique étalée à droite



Scanned with CamScanner

iii) Une distribution asymétrique étalée à gauche



Scanned with CamScanner

Le coefficient d'asymétrie de Fisher (ou skewness) : $\gamma = \frac{\mu_3}{\sigma_X^3}$ où $\mu_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3$ ou $\mu_3 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^3$ est le moment centré d'ordre 3 et σ_X^3 est l'écart-type élevé au cube.

Il est nul si la distribution des fréquences relatives est symétrique autour de la moyenne, négatif si la distribution est unimodale étalée vers la gauche, et positif si la distribution est unimodale étalée vers la droite.

Scanned with CamScanner

3-2- Coefficient d'aplatissement

Le coefficient d'aplatissement Γ de Fisher (ou Kurtosis) : $\Gamma = \frac{\mu_4}{\sigma_X^4} - 3$ où $\mu_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$ ou $\mu_4 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^4$ est le moment centré d'ordre 4 et σ_X^4 est l'écart-type élevé à la puissance 4.

Ce coefficient est nul pour une distribution normale (attention « normale » fait référence à la loi statistique normale), positif ou négatif selon que la distribution est plus ou moins aplatie que la distribution normale de même moyenne et de même écart-type.

Scanned with CamScanner

Exemple 1 (le cas d'une variable discrète):

On reprend l'exemple de la distribution de 100 familles selon le nombre de pièces du logement.

x_i	n_i	$x_i - \bar{x}$	$(x_i - \bar{x})^3$	$n_i(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$	$n_i(x_i - \bar{x})^4$
1	10	-2,86	-23,3937	-233,9366	66,9059	669,0586
2	15	-1,86	-6,4349	-96,5228	11,9688	179,5325
3	15	-0,86	-0,6361	-9,5408	0,5470	8,2051
4	30	0,14	0,0027	0,0823	0,0004	0,0115
5	10	1,14	1,4815	14,8154	1,6890	16,8896
6	12	2,14	9,8003	117,6041	20,9727	251,6728
7	5	3,14	30,9591	154,7957	97,2117	486,0586
8	3	4,14	70,9579	212,8738	293,7659	881,2977
Total	100	-	82,7372	160,1712	493,0614	2492,7264

Scanned with CamScanner

Le coefficient d'asymétrie de Fisher (ou skewness) est :

$\gamma = \frac{\mu_3}{\sigma_X^3} = \frac{160{,}1712/100}{1{,}7721^3} = 0{,}2878 > 0$. D'où la distribution est asymétrique étalée à droite.

Le coefficient d'aplatissement Γ de Fisher (ou Kurtosis) est :

$\Gamma = \frac{\mu_4}{\sigma_X^4} - 3 = \frac{2492{,}7264/100}{1{,}7721^4} - 3 = -0{,}4724 < 0$. La distribution est plus aplatie (une distribution platikurtique) que la distribution normale.

Scanned with CamScanner

Exemple 2 (le cas d'une variable continue):

On reprend l'exemple de la répartition des salaires de 200 salariés de l'entreprise Atlas.

aleurs du salaire n milliers de DH	x_i (le centre de la classe $[e_{i-1}; e_i[$)	n_i	$x_i - \overline{x}$	$(x_i - \overline{x})^3$	$n_i(x_i - \overline{x})^3$	$(x_i - \overline{x})^4$	$n_i(x_i - \overline{x})^4$
$3 \leq x < 6$	4,5	50	-3,9	-59,3190	-2965,9500	231,3441	11567,2050
$6 \leq x < 9$	7,5	90	-0,9	-0,7290	-65,6100	0,6561	59,0490
$9 \leq x < 12$	10,5	30	2,1	9,2610	277,8300	19,4481	583,4430
$12 \leq x < 15$	13,5	15	5,1	132,6510	1989,7650	676,5201	10147,8015
$15 \leq x < 18$	16,5	10	8,1	531,4410	5314,4100	4304,6721	43046,7210
$18 \leq x < 21$	19,5	5	11,1	1367,6310	6838,1550	15180,7041	75903,5205
Total	-	200	-	1980,9360	11388,6000	20413,3446	141307,7400

Scanned with CamScanner

Le coefficient d'asymétrie de Fisher (ou skewness) est :

$\gamma = \frac{\mu_3}{\sigma_X^3} = \frac{11388,6/200}{3,6249^3} = 1,1955 > 0$. D'où la distribution est asymétrique étalée à droite.

Le coefficient d'aplatissement Γ de Fisher (ou Kurtosis) est :

$\Gamma = \frac{\mu_4}{\sigma_X^4} - 3 = \frac{141307,74/100}{3,6249^4} - 3 = 1,0921 > 0$. La distribution est moins aplatie (une distribution leptokurtique) que la distribution normale.

Scanned with CamScanner